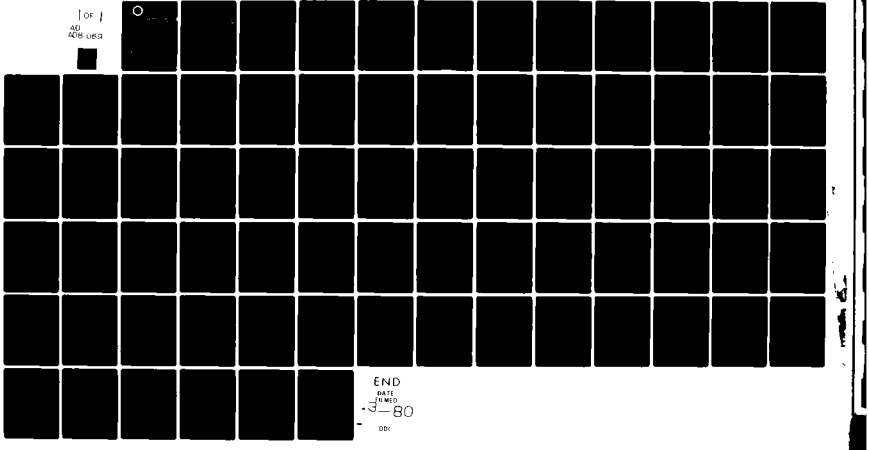
AD-A081 069 STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB F/G 12/1
A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR MINIMAX OPTIMIZA--ETC(U)
NOV 79 W MURRAY, M L OVERTON DAAG29-79-C-0110
UNCLASSIFIED SOL-79-21 ARO-16470.4-M NL



END
DATE FILMED
3-80
DDC

Systems
Optimization
Laboratory

LEVEL II

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
Stanford University
Stanford, California
94305

A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR MINIMAX OPTIMIZATION

by

Walter Murray and Michael L. Overton

TECHNICAL REPORT SOL 79-21
November 1979

74

SOL-79-21

DAAG29-79-C-0110, DE-AS03-76SF00326

ARO 1647.0.4-M

Research and reproduction of this report were partially supported by the Department of Energy Contract ~~DE-AS03-76SF00326,~~ PA Number DE-AT-03-76ER72018; National Science Foundation Grant MCS76-20019 ~~A01,~~ U.S. Army Research Office Contract ~~DAAG29-79-C-0110~~.

A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR MINIMAX OPTIMIZATION

by

Walter Murray

and

Michael L. Overton*

ABSTRACT

The minimax problem is an unconstrained optimization problem whose objective function is not differentiable everywhere, and hence cannot be solved efficiently by standard techniques for unconstrained optimization. It is well known that the problem can be transformed into a nonlinearly constrained optimization problem with one extra variable, where the objective and constraint functions are continuously differentiable. This equivalent problem has special properties which are ignored if solved by a general-purpose constrained optimization method. The algorithm we present exploits the special structure of the equivalent problem. A direction of search is obtained at each iteration of the algorithm by solving an equality-constrained quadratic programming problem, related to one a projected Lagrangian method might use to solve the equivalent constrained optimization problem. Special Lagrange multiplier estimates are used to form an approximation to the Hessian of the Lagrangian function, which appears in the quadratic program. Analytical Hessians, finite-differencing or quasi-Newton updating may be used in the approximation of this matrix. The resulting direction of search is guaranteed to be a descent direction for the mirimax objective function. Under mild conditions the algorithms are locally quadratically convergent if analytical Hessians are used.

1. Introduction.

The problem of concern is

$$\text{MMP: } \min_{\bar{x}} \{F_M(\bar{x}) \mid \bar{x} \in R^n\},$$

$$\text{where } F_M(\bar{x}) = \max \{f_i(\bar{x}),\ i = 1,2,\ldots,m\}$$

and the functions $f_i: R^n \to R^1$ are twice continuously differentiable. The function $F_M(\bar{x})$ is called the minimax function and MMP is usually referred to as the minimax problem. The minimax problem is an unconstrained optimization problem in which the objective function has discontinuous derivatives. Moreover, any solution is usually at a point of discontinuity and consequently it is inappropriate to use any of the known powerful methods for unconstrained minimization to solve MMP. An equivalent problem to MMP is the following nonlinearly constrained problem in which both the objective and constraint functions are twice continuously differentiable:

$$\text{EMP: } \min_{x} \{x_{n+1} \mid x \in R^{n+1}\}$$

$$\text{subject to } c_i(x) \geq 0, \qquad i = 1,2,\ldots,m,$$

$$\text{where } c_i(x) = x_{n+1} - f_i(\bar{x}),\ i = 1,2,\ldots,m,$$

$$\text{and } x^T = (\bar{x}^T, x_{n+1}).$$

We could solve EMP using one of the many methods available for the general

constrained optimization problem:

$$\text{NCP:} \quad \min_{x} \; \{F^{(G)}(x)\}$$

$$\text{subject to} \quad c_i^{(G)}(x) \geq 0, \quad i = 1,2,\ldots,m \ ,$$

where $F^{(G)}$ and $\{c_i^{(G)}\}$ are arbitrary twice continuously differentiable functions. It will be shown, however, that a method can be derived that exploits the special structure of EMP.

The primary special feature of EMP from which many other properties follow is that the minimax function F_M is itself a natural merit function which can be used to measure progress towards the solution of EMP. For problem NCP, in general such a natural merit function is not available, and it is necessary to introduce an artificial one such as a penalty or augmented Lagrangian function to weigh the constraint violation against the decreasing of the objective function, or a barrier function to enforce feasibility. All these merit functions require the definition of a parameter which is to some degree arbitrary and its selection can prove difficult. In the case of penalty and augmented Lagrangian functions, difficulties may also arise because often the global minimum of the merit function is not the solution of the original problem.

The method we adopt to solve MMP essentially consists of two steps at each iteration:

(1) Obtain a direction of search by solving and perhaps modifying an equality-constrained quadratic programming problem (QP), related to one a projected Lagrangian algorithm might use to solve EMP. This procedure is

described in full in subsequent sections.

(2) Take a step along the search direction which reduces the minimax function. Because the minimax function is not differentiable, it is important for efficiency to use a special line search algorithm.

Projected Lagrangian algorithms for solving the general problem NCP via successive quadratic programs have been proposed or analyzed by a number of authors including Wilson (1963), Murray (1969a), Robinson (1974), Wright (1976), Han (1977a), Powell (1977), and Murray and Wright (1978). We make further comments on the extent of the implications of the special structure of EMP, and hence the relationship of our algorithm to these algorithms for the general problem, in Section 15.

A number of other algorithms have been proposed for solving the nonlinear minimax problem. Our approach is most closely related to those due to Han (1977b) and Conn (1979). We will discuss these further in Section 12, after our algorithm has been described in full.

An important special case of MMP is the problem of minimizing the ℓ_∞ norm of a vector function $f(\bar{x}) \in R^m$:

$$\ell_\infty P: \quad \min \{F_\infty(\bar{x}) \mid \bar{x} \in R^n\}$$

$$\text{where} \quad F_\infty(\bar{x}) = \max \{| f_i(\bar{x}) |, \quad i = 1,2,\ldots,m\} .$$

Handling this case in a special manner presents no essential difficulties. However, in order to avoid unnecessarily complicated notation, we postpone

discussion of this until Section 11.

We note that no convexity assumptions are made about the functions $f_i(\bar{x})$. The difficulties of finding global minima without convexity assumptions are well known - - we concern ourselves only with local minima.

1.1 Notation.

Define $\overset{*}{x}$ to be a solution of EMP . It follows that $\overset{*}{\bar{x}}$, the vector composed of the first n elements of $\overset{*}{x}$, is a solution to MMP and $\overset{*}{x}_{n+1} = F_M(\overset{*}{\bar{x}})$.

Let $x^{(k)}$ denote the k-th approximation to $\overset{*}{x}$ and $\bar{x}^{(k)}$ the k-th approximation to $\overset{*}{\bar{x}}$. In general, we will use a $-$ placed above a vector to denote the vector composed of the first n elements of the vector without the $-$.

At each iteration of the algorithm, $x^{(k+1)}$ is obtained by setting

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} + \alpha \bar{p} \quad \text{and} \quad x_{n+1}^{(k+1)} = F_M(\bar{x}^{(k+1)})$$

where $\bar{p}$ is the direction of search and α , a positive scalar, is the steplength. Note that this choice of $x_{n+1}^{(k+1)}$ immediately guarantees that all the points $\{x^{(k)}\}$ are <u>feasible</u> for problem EMP, i.e. $c_i(x^{(k)}) \geq 0$, $i = 1,\ldots,m$.

At any point x we define an <u>active set</u> of constraints of EMP as those which we think will have the value zero at the solution $\overset{*}{x}$, based on the information at x . This set will usually include all constraints with the value zero at the point x and may also include some with positive values. The exact procedure for initially selecting the active set at each iteration

will be discussed in Section 10, and procedures for modifying this choice will be described in Sections 5.2 and 6. We define $t(= t(x))$ to be the number of active constraints at x, and write the vector of active constraints as $\hat{c}(x) \in R^t$. We similarly define $\hat{f}(\bar{x})$ as the vector of active functions corresponding to the active constraints, i.e., those functions expected to have the value $F_M^*(\bar{x})$ at $\bar{x}^*$. Let $\hat{V}(\bar{x})$ be the $n \times t$ matrix whose columns $\{\hat{v}_j(\bar{x})\}$ are the gradients of the active functions, and let $\hat{A}(x)$ be the $(n+1) \times t$ matrix whose columns $\{\hat{a}_j(x)\}$ are the gradients of the active constraints. Thus

$$\hat{A}(x) = \begin{bmatrix} -\hat{V}(\bar{x}) \\ \hat{e}^T \end{bmatrix} \quad \text{where} \quad \hat{e} = (1,\ldots,1)^T \in R^t$$

$$= [\hat{a}_1(x)\ldots\hat{a}_t(x)] = \begin{bmatrix} -\hat{v}_1(\bar{x})\ldots & -\hat{v}_t(\bar{x}) \\ 1 \quad \ldots & 1 \end{bmatrix}$$

We define $Y(x)$ to be a matrix with orthonormal columns spanning the range space of $\hat{A}(x)$, and $Z(x)$ to be a matrix with orthonormal columns spanning the null space of $\hat{A}(x)$. Let I_s be the identity matrix of order s. Provided $\hat{A}(x)$ has full rank, we have that $Y(x)$ has dimension $(n+1) \times t$, $Z(x)$ has dimension $(n+1) \times (n+1-t)$, and

$$Y(x)^T Y(x) = I_t, \quad Z(x)^T Z(x) = I_{n+1-t},$$

$$Y(x)^T Z(x) = \hat{A}(x)^T Z(x) = 0.$$

Let $e_{n+1} = (0,\ldots,0,1)^T \in R^{n+1}$. The Lagrangian function for problem EMP is given by

$$L(x,\lambda) = x_{n+1} - \lambda^T \hat{c}(x) \quad ,$$

where $\lambda \in R^t$ is a vector of Lagrange multipliers. The gradient of $L(x,\lambda)$ with respect to x is $e_{n+1} - \hat{A}\lambda$. We define the $(n+1) \times (n+1)$ matrix $W(x,\lambda)$ to be the Hessian of the Lagrangian function with respect to x. Thus

$$W(x,\lambda) = \sum_{i=1}^{t(x)} -\lambda_i \nabla^2 \hat{c}_i(x)$$

$$= \begin{bmatrix} \overline{W}(\overline{x},\lambda) & 0 \\ 0 & 0 \end{bmatrix}$$

where

$$\overline{W}(\overline{x},\lambda) = \sum_{i=1}^{t(x)} \lambda_i \nabla^2 \hat{f}_i(\overline{x}) \; .$$

The term "projected Hessian of the Lagrangian function" is used to indicate projection into the null space of $\hat{A}(x)$, i.e., the matrix $Z(x)^T W(x,\lambda) Z(x)$. This matrix may also be written $\overline{Z}(x)^T \overline{W}(\overline{x},\lambda) \overline{Z}(x)$, where $\overline{Z}(x)$ consists of the first n rows of $Z(x)$.

Often we will omit the arguments from $\hat{c}$, $\hat{A}$, Z, etc. when it is clear that they are evaluated at $x^{(k)}$. We use the notation $\overset{*}{V}$, $\overset{*}{A}$, $\overset{*}{Z}$, etc. to denote $\hat{V}, \hat{A}, Z$, etc. evaluated at $\overset{*}{x}$ with the active set correctly chosen, i.e., consisting of all those constraints with the value zero at $\overset{*}{x}$.

1.2 Necessary and Sufficient Conditions.

In the following we shall refer to the first- and second-order constraint qualifications and the necessary and sufficient conditions for a point $\overset{*}{x}$ to be a local minimum of the general problem NCP as defined in Fiacco and McCormick (1968). The conditions for $\overset{*}{x}$ to be a local minimum of EMP (and hence $\overset{*}{\bar{x}}$ of MMP) are simplifications of these general conditions. The main simplification is that the first-order constraint qualification always holds for EMP. To see this observe the following. For any point x let p be any nonzero vector satisfying

$$a_i(x)^T p \geq 0 \quad \text{for all } i \text{ s.t. } c_i(x) = 0$$

where the vector a_i is the gradient of c_i. Then p is tangent at $\theta = 0$ to the locally differentiable feasible arc

$$x(\theta) = \begin{bmatrix} \bar{x} + \theta\bar{p} \\ \max \{F_M(\bar{x} + \theta\bar{p}),\ x_{n+1} + \theta p_{n+1}\} \end{bmatrix} .$$

The first-order conditions therefore reduce to the following (see Demyanov and Malozemov (1974) for an alternative derivation applied directly to MMP).

First-order necessary condition.

If $\overset{*}{x}$ is a local minimum of EMP then there exists a vector of Lagrange multipliers $\overset{*}{\lambda} \in R^t$ such that

$$e_{n+1} - \bar{A}\overset{*}{\lambda} = 0 \qquad (1.1)$$

$$\text{and } \overset{*}{\lambda} \geq 0 .$$

Two conditions which are equivalent to (1.1) are that $\overset{*}{x}$ is a stationary point of $L(x,\lambda)$ with respect to x and that $\overset{*T}{Z} e_{n+1} = 0$. Note that (1.1) implies that $\overset{*}{V}$ is rank deficient and that the sum of the multipliers is one.

The second-order constraint qualification does not necessarily hold for EMP (for example at the origin for $f_1 = x_1^3$, $f_2 = -x_1^3$, and $f_3 = -x_1^2$). We therefore include this assumption in the statement of the second-order necessary condition.

Second-order necessary condition.

If $\overset{*}{x}$ is a local minimum of EMP and the second-order constraint qualification holds, then $\overset{*T}{Z}\, W(\overset{*}{x},\overset{*}{\lambda})\overset{*}{Z}$, the projected Hessian of the Lagrangian function, is positive semi-definite.

Sufficient condition.

If the first-order necessary condition holds at $\overset{*}{x}$, the Lagrange multipliers are all strictly positive, i.e. $\overset{*}{\lambda} > 0$, and $\overset{*T}{Z}\, W(\overset{*}{x},\overset{*}{\lambda})\overset{*}{Z}$ is positive definite, then $\overset{*}{x}$ is a strong local minimum of problem EMP. Thus in terms of problem MMP, $F_M(\overset{*}{\bar{x}}) < F_M(\bar{x})$ for all $\bar{x}$ such that $|\bar{x} - \overset{*}{\bar{x}}| < \delta$, for some $\delta > 0$.

Note that in the case where all the f_i are linear it is well known that a solution must exist with $n+1$ active functions at $\overset{*}{\bar{x}}$ (see Cheney (1966) for the case ℓ_∞P). Then normally $\overset{*}{Z}$ is null and therefore the second-order conditions are also null. The nonlinear problem, however, can have a unique solution with anything from 1 to $n+1$ functions active at $\overset{*}{\bar{x}}$. This relationship is exactly analogous to that between linear and nonlinear programming. For comments on the special case of ℓ_∞ approximation and the meaning of the Haar condition, see Section 11.

2. Use of the Equivalent Problem EMP.

Clearly it is desirable that at every iteration the search direction $\bar{p}$ be a descent direction for F_M, i.e.

$F'_M(\bar{x}^{(k)}, \bar{p}) < 0$, where $F'_M(\bar{x}^{(k)}, \bar{p})$ is the directional derivative $\lim_{h \to 0^+} \frac{1}{h} (F(\bar{x}^{(k)} + h\bar{p}) - F(\bar{x}^{(k)}))$.

An equivalent condition is that $\bar{p}$ is a descent direction for each function f_i for which $f_i(\bar{x}^{(k)}) = F_M(\bar{x}^{(k)})$ (i.e. $c_i(x^{(k)}) = 0$) . A second desirable property for $\bar{p}$ arises from considering the active set which consists of those constraints corresponding to functions we expect to have the value $F_M(\bar{x}^*)$ at $\bar{x}^*$. We wish to choose $\bar{p}$ so that the first order change in these functions predicts that they will all have the same value at $\bar{x}^{(k)} + \bar{p}$. An equivalent condition is:

$$\hat{A}(x^{(k)})^T p = - \hat{c}(x^{(k)}) \tag{2.1}$$

and hence

$$\hat{f}_i(\bar{x}^{(k)}) + \hat{v}_i(\bar{x}^{(k)})^T \bar{p} = F_M(\bar{x}^{(k)}) + p_{n+1}, \quad i = 1,\ldots,t$$

for some value p_{n+1} (the (n+1)st component of p). If the active set included all the constraints which are zero at $x^{(k)}$, then the condition

$$p_{n+1} < 0 \tag{2.2}$$

also ensures that $\bar{p}$ is a descent direction for F_M . In fact, at every iteration the active set will initially include all such constraints, but it may be desirable to drop one or more of them from the set to move

off a constraint. Since the decrease in F_M is limited to the smallest decrease in any of the functions corresponding to $c_i(x^{(k)}) = 0$, it is also desirable to insist that

$$a_i(x^{(k)})^T p \geq 0 \text{ for all } i \text{ such that } c_i(x^{(k)}) = 0 \ . \qquad (2.3)$$

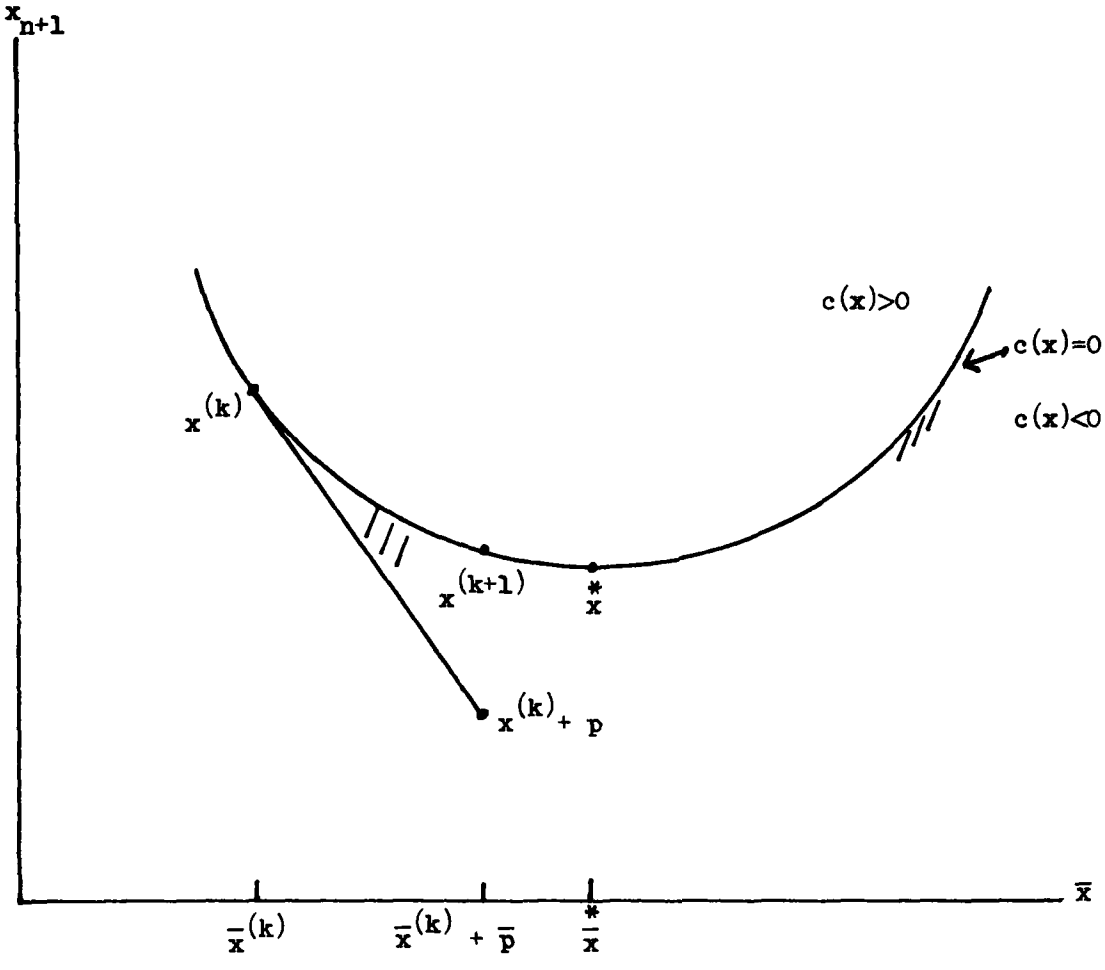
Conditions (2.2) and (2.3) ensure that p is a <u>first-order feasible descent direction</u> with respect to problem EMP. It is straightforward to show the following from the above remarks.

<u>Theorem 1.</u> If (2.2) and (2.3) hold, i.e., p is a first-order feasible descent direction for EMP, then $\bar{p}$ is a descent direction for F_M and hence a sufficiently small step along it must result in a reduction in F_M.

Note that the above does not guarantee that p is a feasible direction for EMP, as illustrated by Figure 1. This causes no difficulty since $x_{n+1}^{(k+1)}$ is set to $F(\bar{x}^{(k)})$ and hence it is always possible to obtain a lower feasible point for EMP if (2.2) and (2.3) hold. Consequently it is important to look for a reduction in F_M in the line search and not in a penalty function constructed for EMP. In Figure 1,

$$F_M(x_1^{(k)} + p_1) < F_M(x_1^{(k)})$$

but a penalty function may have a much higher value at $x^{(k)} + p$ than at $x^{(k)}$ since $x^{(k)} + p$ is infeasible for EMP.



Example for n=m=1 where p is first-order feasible but not feasible for EMP.

FIGURE 1 .

Thus we see that the importance of EMP is as a device to obtain a search direction $\bar{p}$ along which F_M can be reduced in the line search. We emphasize again that we wish (2.2) and (2.3) to hold so that $\bar{p}$ is a descent direction for F_M, and that the active set nature of the algorithm indicates that (2.1) should also hold. It is not reasonable to carry (2.1) one step further and demand that the second order change in the active functions predict they all have the same new value, since this would require computing and storing each of the Hessians of the active functions, a process which is unwarranted especially since it may not be possible to solve the resulting equations for p. It is the case however that (2.1), (2.2) and (2.3) will usually not uniquely define $\bar{p}$, and in the next section we utilize properties of EMP to obtain an initial choice of $\bar{p}$ by solving a quadratic program (QP) based on second order information incorporated in an approximation to the Lagrangian function. The solution to this QP may not always satisfy (2.1), (2.2) and (2.3), and in subsequent sections we discuss how to modify the initial choice to obtain a satisfactory search direction.

3. The QP Subproblem

The solution of EMP is at a minimum of the Lagrangian function in the null space of the active constraint Jacobian at $\overset{*}{x}$.

The usual method for solving a general linearly constrained problem is to approximate the objective function by a quadratic function and

then determine the search direction by solving some appropriate quadratic program (QP). Consider therefore the quadratic program

$$\min_{p} \; L(x^{(k)}, \lambda^{(k)}) + (e_{n+1} - \hat{A}(x^{(k)})\lambda^{(k)})^T p + \frac{1}{2} p^T W(x^{(k)}, \lambda^{(k)}) p$$

$$\text{subject to} \quad \hat{A}(x^{(k)})^T p = - \hat{c}(x^{(k)}) .$$

where $\lambda^{(k)}$ is an approximation to $\overset{*}{\lambda}$.

An equivalent QP is given by

$$\text{QP1} : \min_{p} \; \frac{1}{2} p^T W(x^{(k)}, \lambda^{(k)}) p + e_{n+1}^T p$$

$$\text{subject to} \quad \hat{A}(x^{(k)})^T p = - \hat{c}(x^{(k)}) .$$

Let us drop the arguments $x^{(k)}$ and $\lambda^{(k)}$ from $\hat{A}$ and W, and let Y and Z be the orthogonal matrices defined in Section 1.1. The matrices Y and Z may be determined from the QR factorization of $\hat{A}$:

$$\hat{A} = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = [Y \;\; Z] \begin{bmatrix} R \\ 0 \end{bmatrix} ,$$

where R is an upper triangular matrix of order t. If $\hat{A}$ has full rank and Z^TWZ is positive definite, then the unique solution of QP1 can be expressed as the sum of two orthogonal components:

$$p = Yp_Y + Zp_Z \tag{3.1}$$

where $p_Y \in R^t$ and $p_Z \in R^{n+1-t}$.

We have

$$\hat{A}^T p = R^T p_Y = -\hat{c} \tag{3.2}$$

and p_Y is determined entirely by the constraints of QP1.
The vector p_Z is given by the solution of

$$(Z^TWZ)p_Z = -Z^T(e_{n+1} + WYp_Y) \tag{3.3}$$

(see Murray and Wright (1978)).

In subsequent sections we will also wish to refer to a related QP and its solution, namely the one with the same quadratic form but homogeneous constraints:

$$\text{QP2}: \quad \min_p \quad \frac{1}{2} p^T Wp + e_{n+1}{}^T p$$
$$\text{subject to} \quad \hat{A}(x^{(k)})^T p = 0 .$$

The solution to this is given by $p = Zq_Z$, where

$$(Z^TWZ)q_Z = -Z^T e_{n+1} . \tag{3.4}$$

At every iteration of our algorithm an attempt is made to set the search direction p to the solution of QP1, but for various reasons this may be inadequate (there may not even be a solution). Much of the detailed discussion of the method is concerned with what action to take in these circumstances. It is important to realize that it is only on attempting to solve the QP that these inadequacies are revealed and if the solution is not sought in a particular manner, certain deficiencies are not ascertained. We are restricted in the action we may take by requiring that the search direction satisfy (2.1), (2.2) and (2.3), and by a need to limit the computational effort when the information on which it is based has proved suspect. We also wish to arrange the computation so that any computational efforts already invested can still be utilized should the initial QP prove inadequate. In particular, provision must be made for the possibility that the wrong active set is identified at $x^{(k)}$. We will always insist that at the beginning of every iteration the potential active set includes all constraints with zero value at $x^{(k)}$ (and it will normally also include constraints with positive values). In Section 5.2 and 6 we discuss how a constraint may then, if desired, be deleted from the active set.

For the moment we assume that analytical Hessians are used to compute W, but in Section 7 we discuss finite difference and quasi-Newton alternatives. Notice that the definition of W requires an estimate of the Lagrange multipliers λ^* to be made available at $x^{(k)}$. Before discussing these other matters, we devote the next section to the subject of multiplier estimates.

4. Lagrange Multiplier Estimates.

Although the arguments of Section 3 amply justify defining the basic search direction p as the solution to QP1, it is not nearly as clear how to define the Lagrange multiplier estimate at $x^{(k)}$. These estimates are needed to define the matrix W and to determine whether constraints should be deleted from the active set. Clearly it is better to use new information obtained at the current point $x^{(k)}$ rather than use the multipliers of the QP solved at the previous iteration.

λ_L: The least squares estimate.

The most obvious multiplier estimate is the least squares solution to the overdetermined system based on the first order necessary conditions. Thus we define λ_L to be the solution of the least squares problem

$$\min_{\lambda} \quad \|\hat{A}\lambda - e_{n+1}\|_2^2 \quad .$$

It is well-known (see Golub (1965)) that the least squares problem can be solved by using the QR factorization of $\hat{A}$, which we introduced in the last section to solve QP1. The estimate λ_L is called a first-order multiplier estimate because

$$\|\lambda_L - \lambda_M\| = O(\|x^{(k)} - x_M\|)$$

where x_M is the minimum of EMP on the manifold defined by the current active set, i.e., the solution of

$$\min_{x} \quad x_{n+1} \qquad \text{subject to} \quad \hat{c}(x) = 0$$

and λ_M is the corresponding Lagrange multiplier vector.

The multiplier estimate λ_L has the following property. Suppose for some reason p is unsatisfactory and we wish to delete a constraint from the active set. If $(\lambda_L)_j < 0$ and we delete constraint j, then the steepest descent direction in the nullspace of the new active constraint Jacobian is guaranteed to be first-order feasible with respect to the deleted constraint. Define $\tilde{A}$ as $\hat{A}$ with $\hat{a}_j$ deleted, and $\tilde{Z}$ by

$$\tilde{A}^T\tilde{Z} = 0 , \quad \tilde{Z}^T\tilde{Z} = I_{n+2-t}, \quad \tilde{Z} = [Z \quad z] . \tag{4.1}$$

Then the steepest descent step in the new null space is given by

$$\tilde{Z}\, s_{\tilde{Z}} = - \tilde{Z}\, \tilde{Z}^T e_{n+1} \tag{4.2}$$

and we have

$$\hat{a}_j^T \tilde{Z}\, s_{\tilde{Z}} > 0 \tag{4.3}$$

The Newton step in the null space, given by $\tilde{Z}\, q_{\tilde{Z}}$, where

$$(\tilde{Z}^T W \tilde{Z}) q_{\tilde{Z}} = - \tilde{Z}^T e_{n+1} \tag{4.4}$$

does not in general satisfy

$$\hat{a}_j^T \tilde{Z}\, q_{\tilde{Z}} \geq 0 . \tag{4.5}$$

See Gill and Murray (1979) for the proof of these statements in the context of linearly constrained optimization.

λ_C: A special estimate for problem EMP.

A more appropriate estimate than λ_L can be obtained by considering the special structure of EMP. The least squares solution

is motivated by the fact that the overdetermined set of equations $\hat{A}\lambda = e_{n+1}$ is only an approximation to the set of equations which hold at x_M, the minimum on the manifold. However, the $(n+1)$st equation $\hat{e}^T\lambda = 1$ is exactly, not approximately, the equation which holds at x_M. We therefore define another multiplier estimate λ_C as the least squares solution to the first n equations subject to the constraint $\hat{e}^T\lambda_C = 1$. Thus λ_C is the solution to the constrained least squares problem

$$\min_{\lambda} \quad \|\hat{V}\lambda\|_2^2 \qquad \text{subject to} \quad \hat{e}^T\lambda = 1 \quad .$$

In fact we can show that λ_C is exactly λ_L multiplied by a scalar greater than or equal to one.

Theorem 2. Assume $\hat{A}$ has full rank and let λ_L and λ_C be defined as above. Then $\lambda_C = \beta\lambda_L$, where $\beta = \dfrac{1}{\hat{e}^T\lambda_L} \geq 1$.

Proof. If $\hat{V}$ does not have full rank then let $\lambda \neq 0$ be such that $\hat{V}\lambda = 0$. We have $\hat{e}^T\lambda \neq 0$ since otherwise $\hat{A}\lambda = 0$. Thus $\lambda_C = \lambda_L = \dfrac{1}{\hat{e}^T\lambda}\lambda$ with both the least squares and the constrained least squares problems having zero residuals, and the result follows. Therefore assume that $\hat{V}$ has full rank. The vector λ_L satisfies

$$\lambda_L = (\hat{A}^T\hat{A})^{-1}\hat{A}^T e_{n+1}$$

$$= (\hat{V}^T\hat{V} + \hat{e}\hat{e}^T)^{-1}\hat{e} \quad .$$

It follows from the Sherman-Morrison formula (see Householder (1964, p. 123)), that

$$\lambda_L = d - \frac{1}{1+\hat{e}^T d}\, d\, \hat{e}^T d$$

$$= \frac{1}{1+\hat{e}^T d}\, d \tag{4.6}$$

where

$$d = (\hat{V}^T\hat{V})^{-1}\hat{e} \tag{4.7}$$

Note that $\hat{e}^T d > 0$ since $(\hat{V}^T\hat{V})^{-1}$ is positive definite.

Since λ_C is the solution to a constrained optimization problem it follows from the corresponding first order necessary conditions (writing $\|\hat{V}\lambda\|^2 = \lambda^T\hat{V}^T\hat{V}\lambda$) that

$$2\,\hat{V}^T\hat{V}\,\lambda_C = \beta_1\hat{e}$$

where β_1 is the Lagrange multiplier associated with the constraint $\hat{e}^T\lambda = 1$. Since $\hat{e}^T d > 0$ and $\hat{e}^T\lambda_C = 1$ we have

$$\lambda_C = \frac{1}{\hat{e}^T d}\, d$$

and hence the result follows. Alternatively the equation for λ_C could have been obtained by scaling the last equation in the unconstrained least squares problem and observing that the right hand side of (4.6), modified to include the scale factor, tends to $\frac{1}{\hat{e}^T d}\, d$ as the scale factor goes to infinity. □

Note that it would be highly ill-advised to compute λ_L and λ_C by computing d via (4.7). If $x^{(k)}$ were equal to x_M, the minimum on the manifold, then $\hat{V}$ would be rank-deficient, but $\hat{A}$ would not

in general, and hence if $x^{(k)}$ is close to x_M the condition number of $\hat{V}^T\hat{V}$ may be much bigger than that of $\hat{A}^T\hat{A}$. Since solving the least squares problem using the QR factorization of $\hat{A}$ is alrady a better conditioned process than explicitly using $\hat{A}^T\hat{A}$ (via the normal equations) it is clear that using the QR factorization of $\hat{A}$ is far preferable to using (4.7).

Using the scaled estimate λ_C instead of λ_L will result in different decisions about whether to delete constraints from the active set since, as will be explained in Section 6, the magnitudes as well as the signs of the estimates are used to make the decision. Furthermore, using λ_C instead of λ_L in general increases the magnitude of W and hence affects both the direction (if $\hat{c} \neq 0$) and the magnitude of the solution to QP1. The following example illustrates that it may often be beneficial to use λ_L rather than λ_C. Let $n = m = 1$ and $F(\bar{x}) = f_1(\bar{x}) = \bar{x}^2$. Let the current estimate of the solution be $\bar{x}^{(k)} = 2$. Then $\hat{V} = 4$ and $\hat{A} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$, $\lambda_L = \frac{1}{17}$ and $\lambda_C = 1$. Using λ_C for W results in the exact step to the solution being taken, but using λ_L results in one seventeen times too big. Clearly similar examples can be constructed with a larger number of active constraints. The choice of λ_C over λ_L essentially arises from the fact that problem MMP is in some sense naturally scaled -- the functions of MMP cannot be individually scaled without changing the solution while the constraints of NCP can be individually scaled in general.

λ_D: A second special estimate.

The estimate λ_C is not the only multiplier estimate that might be constructed to take the special form of EMP into account. Since we know that $\hat{V}$ would be rank-deficient if $x^{(k)}$ were the minimum on the manifold, a reasonable alternative would be to assume that $\hat{V}$ is close to being rank-deficient and define λ_D as the solution to an appropriate rank deficient least squares problem, scaled so that $\hat{e}^T\lambda_D = 1$. Let the QR factorization of $\hat{V}$ with column pivoting be

$$\hat{V} = Q_1 \begin{bmatrix} R_1 & s \\ & \tau \end{bmatrix} \Pi$$

where Q_1 is an orthogonal matrix, Π is a permutation matrix, s is a vector and τ is a scalar. If $\tau = 0$ then the vector

$$\lambda' = \Pi \begin{bmatrix} -R_1^{-1}s \\ 1 \end{bmatrix}$$

and its multiples are solutions to the rank-deficient least squares problem $\hat{V}\lambda \approx 0$. Therefore let

$$\lambda_D = \frac{1}{\hat{e}^T\lambda'} \lambda' \quad .$$

The estimate λ_D is not a multiple of λ_L in general and if a constraint is deleted corresponding to $(\lambda_D)_j < 0$, it is not true that the steepest descent step satisfies (4.3). This alone makes the use of λ_D questionable. Furthermore, to obtain λ_D it would be necessary that the QR factorization of $\hat{V}$ be done with column pivoting (otherwise it makes no sense to ignore τ) and this means it could not be obtained

by simply updating the factorization of $\hat{A}$. The main flaw of λ_D is that it ignores the information given by τ even though $\hat{V}$ may not be nearly rank deficient. For these reasons we do not consider λ_D any further.

μ_W : A second-order estimate.

The second-order Lagrange multiplier estimate μ_W is defined as the exact multipliers corresponding to the solution p of QP1, i.e. the solution to the consistent set of equations

$$\hat{A}\mu_W = e_{n+1} + Wp \quad , \tag{4.7}$$

where p is given by (3.1), (3.2) and (3.3) and a first-order multiplier estimate is used to define W. The necessity of requiring W to define μ_W implies that second-order estimates can only be useful in determining whether to delete constraints from the active set. The estimate is called second-order because

$$\|\mu_W - \lambda_M\| = O(\|x^{(k)} - x_M\|^2) .$$

If a constraint is deleted corresponding to $(\mu_W)_j < 0$, then (4.3) will not hold in general. If $\hat{c} = 0$, then (4.5) will hold.

The system (4.7) is consistent because of the definition of p. Note that because it is consistent there is no question of a second-order estimate analogous to λ_C - - the last equation is already satisfied by μ_W.

ν_W : A first-order estimate guaranteeing a first-order feasible Newton step after deletion.

The multiplier estimates λ_L and μ_W are often used in constrained optimization and are discussed in some detail in Gill and Murray (1979). Here we note that there is yet another estimate which might be useful both for EMP and in the context of general constrained optimization. We denote it ν_W and define it by

$$\hat{A}\nu_W = e_{n+1} + WZq_Z \quad ,$$

where Zq_Z is the solution to QP2 as defined by (3.4). The estimate ν_W is the vector of exact multipliers corresponding to the solution of QP2. Unlike the case with λ_L or μ_W, if a constraint is deleted corresponding to a negative component of ν_W, then (4.5) must hold even if $\hat{c} \neq 0$, i.e. the Newton step in the new null space must be first-order feasible w.r.t. the deleted constraint. This fact follows from the proof of Theorem 4 in Gill and Murray (1979), where (4.5) is shown to hold in the context of linear constraints. The proof is applicable to ν_W but not μ_W in the context of nonlinear constraints, since ν_W is the exact multiplier vector for a QP with homogeneous constraints. Notice however that unlike μ_W, ν_W is only a first-order estimate.

Using the estimate λ_C to define W.

Both the estimates λ_C and μ_W will be used to decide when to delete constraints from the active set, as will be discussed in Section 6. As explained there, a constraint with a negative component of λ_C will not

necessarily be deleted from the active set, since the multiplier estimates may not be reliable. However, we also use λ_C to define W and there is no good reason to include in W a term with a negative component of λ_C. Therefore, we define W to be

$$W(x^{(k)}) = \sum_{i=1}^{t} \lambda_i' \nabla^2 \hat{f}_i(x^{(k)})$$

where

$$\lambda' = \frac{1}{\hat{e}^T \lambda''} \lambda''$$

and λ'' is defined by $\lambda_i'' = \max(0, (\lambda_C)_i)$.

5. Properties of Solution of QP Subproblem.

In this section we examine the properties of the solution to QP1. Initially we assume that all constraints with zero value are included in the active set and that $\hat{A}$ has full rank and Z^TWZ is positive definite so that the solution p is given by (3.1), (3.2) and (3.3) and is unique. We would like p to satisfy (2.1), (2.2) and (2.3). Clearly the constraints of QP1 ensure that (2.1) and (2.3) hold. Thus the only question is whether p is a descent direction for EMP, i.e. whether (2.2) holds. If all the active constraints have the value zero then the following applies:

Theorem 3. Suppose that $\hat{c} = 0$, $\hat{A}$ has full rank and Z^TWZ is positive definite. Then p, the solution of QP1, is a descent direction for EMP provided it is not zero.

Proof. Since $\hat{A}$ has full rank and the columns of Y span the range of of the columns of $\hat{A}$, we have $p_Y = 0$. Hence $p = Zp_Z$ and

$$e_{n+1}^T p = p_Z^T Z^T e_{n+1}$$

$$= - p_Z^T Z^T W Z p_Z \qquad \text{by (3.3)}$$

Since Z^TWZ is positive definite, $e_{n+1}^T p$ must be negative if $p_Z \neq 0$, i.e. $p \neq 0$. □

If $p = 0$, then by (3.3) $Z^T e_{n+1} = 0$ and hence $\hat{A}\lambda_L = e_{n+1}$ is a consistent set of equations, with $\lambda_C = \lambda_L = \mu_W = \lambda_M$. Thus either one of the components of λ_C is negative or zero, or the first and second order sufficiency conditions are satisfied and $x^{(k)}$ is a solution

of problem EMP. If $p = 0$ and at least one of the multipliers is negative, then it is necessary to delete a corresponding constraint from the active set to obtain a descent direction. The procedure for doing this is described in Section 6. If $p = 0$ and the smallest component of λ_C is zero, the point $x^{(k)}$ may or may not be a solution. In this case special techniques such as described in Gill and Murray (1977) must be used to determine whether to treat the corresponing constraint as active or not. These will not be discussed any further here.

In the next three subsections we discuss the actions which it may be necessary to take to obtain a search direction which satisfies (2.1), (2.2) and (2.3) when we drop the assumptions that $\hat{c} = 0$, $\hat{A}$ has full rank and $Z^T W Z$ is positive definite.

5.1 Positive Active Constraints.

In practice it will rarely be the case that $\hat{c} = 0$, so we now drop this assumption. We note that if we were sufficiently restrictive in the definition of the active set (e.g. choose only one constraint active) then we could force this condition to be true. As will be shown in Section 10,however, it is important for the efficiency of the algorithm not to be too restrictive in the definition of the active set. This may appear to negate the significance of Theorem 3, but this is not the case. Although dropping the assumption certainly means that Theorem 3 no longer holds,since both Yp_Y and Zp_Z could be ascent directions,we can infer that if $\|\hat{c}\|$ is small (and it approaches zero near the solution), then p is likely to be a descent direction. If,having solved QP1,it transpires

that $e_{n+1}^T p > 0$, then, provided either Yp_Y or Zp_Z is a descent direction, a suitable descent direction can be chosen as follows:

$$Yp_Y + \gamma\, Zp_Z \qquad \text{if } e_{n+1}^T Yp_Y < 0$$

$$\gamma\, Yp_Y + Zp_Z \qquad \text{if } e_{n+1}^T Zp_Z < 0$$

for some γ satisfying $0 < \gamma \leq 1$. If neither component is a descent direction then a further possibility is to replace Zp_Z by the solution of QP2, i.e. Zq_Z where q_Z is given by (3.4). Provided $Z^T e_{n+1} \neq 0$, the vector Zq_Z is a descent direction, and hence so is $\gamma Yp_Y + Zq_Z$ for some γ, $0 < \gamma \leq 1$.

When $Z^T e_{n+1} = 0$ and Yp_Y is an descent direction, it is necessary to delete a constraint to obtain a descent direction. One possibility would be to delete constraint j (say) where $\hat{c}_j$ is the largest component of $\hat{c}$, since it is not strictly necessary to be concerned about whether the resulting search direction has a positive inner product with $\hat{a}_j$ ((2.3) applies only to constraints with value zero). However, clearly this may lead to a very small step being taken along the search direction with this constraint being forced immediately back into the active set. The following result shows that when Yp_Y is uphill, a constraint can always be found with a negative multiplier estimate and hence can be deleted more safely. The result is also useful when $Z^T e_{n+1} \neq 0$ since then the fact that Yp_Y is uphill implies there are too many constraints in the active set.

Theorem 4. Assume $\hat{A}$ has full rank and let Yp_Y be defined by (3.2).

If $e_{n+1}{}^T Y p_Y > 0$, then one of the components of λ_C is negative.

Proof. We know λ_C is a positive multiple of λ_L. The vector λ_L satisfies

$$\hat{A}\,\lambda_L = YY^T e_{n+1} \tag{5.1}$$

This is a characterization of the least squares solution using the projector matrix YY^T (see Stewart(1973, p. 228)). Thus

$$\lambda_L{}^T \hat{A}^T Y p_Y = e_{n+1}{}^T YY^T Y p_Y$$

and hence

$$-\lambda_L{}^T \hat{c} = e_{n+1}{}^T Y p_Y > 0 \quad .$$

Since $\hat{c} \geq 0$ it follows that at least one of the components of λ_L and hence λ_C must be negative. □

It also follows from the above proof that if $e_{n+1}{}^T Y p_Y = 0$, then either $\hat{c} = 0$ (covered by Theorem 3), or the minimum element of λ_C is zero or negative.

It is worth noting that Theorem 4 does not hold in general if other multiplier estimates such as λ_D and μ_W are substituted for λ_C.

It follows from Theorem 4 that if Yp_Y is an ascent direction, we can delete the constraint corresponding to a negative component of λ_C to obtain a first-order feasible descent direction. The procedure for doing this is described in Section 6.

5.2 Avoiding Rank Deficiency in the Active Contraint Jacobian.

In this section we demonstrate how the active set can always be

chosen so as to avoid rank deficiency in $\hat{A}$. We first consider the consequences of $\hat{A}$ being rank deficient. If $\hat{A}$ is rank deficient and $\hat{c} \neq 0$, it is not possible in general to satisfy the constraints of QP1 since they may not be compatible. In such circumstances it might be thought that an adequate compromise would be the least squares solution to $\hat{A}^T p \approx - \hat{c}$, but this may not be first-order feasible with respect to EMP and hence may not be a descent direction for the minimax function even if it is a descent direction for EMP. Clearly it is desirable to restrict the number of constraints in the active set so that $\hat{A}$ has full rank.

The way this is done is as follows. Given a set of candidates for the active set, we determine which are to be actually included in the active set during the QR factorization of $\hat{A}$. Assuming the problem is well-scaled, a reasonable order in which to consider the candidates for inclusion is given by the size of the constraint values. Therefore we order the potential columns of $\hat{A}$ by increasing size of $\{c_j\}$ before proceeding to do a QR factorization of the matrix by columns without column pivoting (except where several columns correspond to the same magnitude of c_j). If it transpires during the factorization that any of the potential columns of $\hat{A}$ is linearly dependent on those already included, then the corresponding constraint is not included in the active set. Clearly such a process results in a matrix $\hat{A}$ that has full column rank.

An example which illustrates this procedure is the following. Suppose the initial candidates for the active set are

$$c = \begin{bmatrix} 0 \\ 10^{-4} \\ 10^{-3} \\ 10^{-2} \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 0.5 & 0 \\ 0 & 0 & 0.5 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} .$$

Then the first and third constraints are selected for the active set, and the second and fourth are ignored.

We must now show that omitting any constraint, say $c_r(x)$, from the active set to avoid rank deficiency in $\hat{A}$ does not cause (2.3) to be violated. Strictly speaking, we need not be concerned with constraints for which $c_r(x) > 0$, but for the moment we will not assume $c_r(x) = 0$ since the following also serves to illustrate why we put potentially active constraints in increasing order.

Let a_r be the gradient of c_r. We have

$$a_r = \sum_{i=1}^{s} w_i \hat{a}_i \tag{5.2}$$

for some index s and scalars w_i, $i = 1,\ldots,s$. Since the constraints were ordered we also have

$$c_r \geq \hat{c}_s \geq \hat{c}_{s-1} \geq \cdots \geq \hat{c}_1 \tag{5.3}$$

It follows from (3.2) and (5.2) that

$$a_r^T p = -\sum_{i=1}^{s} w_i \hat{c}_i \tag{5.4}$$

Since we have no a priori information about how the sizes of the $\{w_i\}$ would change if the columns were ordered differently, we have by putting the constraints in increasing order attempted to prevent $|a_r^T p|$ from being large and in particular to prevent $a_r^T p$ from being a large negative number. Ordering the constraints also ensures that the omitted constraint c_r have as large a value as possible, and these two facts combine to make it as unlikely as possible that constraint r will prevent a significant step from being taken along p. It follows from (5.3) and (5.4) that if $c_r = 0$, then $a_r^T p = 0$ and hence (2.3) is satisfied. However,if it is necessary to delete a further constraint from the active set as described in Section 6, then (2.3) may no longer hold. This difficulty is circumvented by allowing a zero step to be taken along p (see Section 8) and then reconsidering the choice of active set in the next iteration.

5.3 Projected Hessian not Positive Definite.

If $Z^T W Z$ is not positive definite, it makes no sense to use the solution of (3.3), even if it exists, to define the direction of search. Such a direction is a step to a stationary point which is not a minimum of the quadratic form on the subspace defined by the constraints and may even be a maximum. It also makes no sense to reduce the number of active constraints (the projected Hessian will still be indefinite) since one means of ensuring that the projected Hessian is positive definite is to increase the number of active constraints. An alternative definition of the search direction which is satisfactory is to utilize the modified Cholesky algorithm of Gill and Murray (1974).

The following numerically stable matrix factorization is computed:

$$Z^T W Z + E = L D L^T$$

The matrix E is a non-negative diagonal matrix large enough to make $Z^T W Z + E$ numerically positive definite. The modified Newton direction in the null space of $\hat{A}$ is then defined as Zq_Z where

$$L D L^T q_Z = - Z^T e_{n+1} \quad . \tag{5.5}$$

Recently several more complicated methods have been suggested for handling indefinite Hessians ; see Fletcher and Freeman (1977), and Moré and Sorenson (1979). All of these, including the Gill and Murray algorithm, provide ways of obtaining directions of negative curvature. As will be explained shortly, however,these are not so useful in the context of nonlinear constraints;our main concern here is simply to obtain a reasonable descent direction.

Like the Newton direction Zq_Z in the positive definite case, the vector Zq_Z is a descent direction provided that $Z^T e_{n+1} \neq 0$. If $Z^T e_{n+1} = 0$ and $e_{n+1}^T Y p_Y < 0$ we can simply set $p = Yp_Y$; if $Z^T e_{n+1} = 0$ and $e_{n+1}^T Y p_Y > 0$ we have already explained that a constraint may be deleted. Similarly if $e_{n+1}^T Y p_Y = 0$, but a component of λ_C is negative, a constraint may be deleted.

If $Z^T W Z$ is not positive semi-definite and $e_{n+1}^T Y p_Y = 0$, $Z^T e_{n+1} = 0$ and $\lambda_C \geq 0$, then the point $x^{(k)}$ is a constrained saddle point (or even a maximum). It is desirable to compute a feasible arc of negative curvature with respect to $Z^T W Z$, which must exist. For the case of linearly constrained minimization of a differentiable function,

Gill and Murray (1974) have shown how to compute a vector of negative curvature using the modified Cholesky factorization. However, it is not possible in general to compute such an arc for the nonlinearly constrained case given only W -- it is necessary to know all the individual constraint Hessians. Such a computation would hence require an unreasonable amount of storage as well as computational effort. We therefore suggest the following. If $\hat{c}_j > 0$ for some j, delete the j-th active constraint and avoid, or at least postpone, the problem. If $\hat{c} = 0$, compute the direction of negative curvature assuming the constraints are linear, which is thus feasible to first order, and try stepping along it. If the minimax function is lower at the new point,then take this as $x^{(k+1)}$. Otherwise an alternative is to try to obtain a lower point using a function-comparison method such as described in Swann (1972). There is no reasonable procedure which is guaranteed to work in this situation. However, it should be noted that such a situation is unlikely to occur since the basic modified Newton iteration of solving successive quadratic programs seeks to avoid saddle points and maxima.

6. Deleting Constraints from the Active Set.

In Section 5 we explained that when $Z^T e_{n+1} = 0$ and a multiplier is negative, it is necessary to delete a constraint from the active set in order to obtain a first-order feasible descent direction. It is well known in the context of constrained optimization that it is ill-advised to wait until $Z^T e_{n+1}$ is zero or very close to zero before deleting a constraint corresponding to a negative multiplier estimate, since doing so is solving the wrong equality-constrained subproblem with unnecessary

accuracy (minimizing on a manifold). It is also well known that it is even more ill-advised to delete a constraint too early, when the multiplier estimates are not yet reliable, since this may cause the constraint to be repeatedly added to and dropped from the active set. Gill and Murray (1979) suggest computing both λ_L and μ_W when possible, never considering deleting a constraint unless the estimates have the same sign and agree within a certain tolerance. Here we use λ_C instead of λ_L. Notice that $\lambda_C = \mu_W$ when $Z^T e_{n+1} = 0$. A further possibility is to insist that the following hold before deleting a constraint:

$$\|Z^T e_{n+1}\| < \delta \cdot \min (1, - \min_i (\lambda_C)_i)$$

where $\delta < 1$ is a constant, say $\delta = 1$. In effect, this test ensures that the higher the uncertainty that a multiplier is negative, the greater the accuracy to which the minimum is approximated on the manifold.

In Section 5.1 we also pointed out a further situation in which we delete a constraint. This is when $e_{n+1}^T Y p_Y > 0$, which means (as shown in Theorem 4) that $(\lambda_C)_j < 0$ (and $\hat{c}_j > 0$) for some j. A constraint is always deleted in this situation since this is a clear indication that there are too many constraints in the active set.

In the rest of this section we discuss what search direction p to choose after deleting constraint j with $(\lambda_C)_j < 0$. Define $\tilde{A}$, $\tilde{Z}$ and z by (4.1). Define $\tilde{c}$ to be $\hat{c}$ with $\hat{c}_j$ deleted. Since $\tilde{A}$ corresponds to the new active set we wish p to satisfy

$$\tilde{A}^T p = - \tilde{c} \qquad (6.1)$$

corresponding to (2.1), and also $p_{n+1} < 0$, i.e. (2.2). If $\hat{c}_j = 0$ then we must have

$$\hat{a}_j^T p \geq 0 \tag{6.2}$$

to satisfy (2.3), but as explained in Section 5.1, this is desirable even if $\hat{c}_j > 0$.

As we noted in Section 4, the steepest descent step $\tilde{Z}s_{\tilde{Z}}$ in the null space of the new set of constraints, given by (4.2), must satisfy (6.2), but the Newton step given by (4.4) may not satisfy (6.2). A step satisfying (6.2) which is preferable to the steepest descent step is a combination of the Newton step in the null space of $\hat{A}$, i.e. Zq_Z, and the steepest descent step in the new direction permitted by deleting the constraint.

Theorem 5. Assume $\hat{A}$ has full rank.
Suppose $(\lambda_C)_j < 0$. Define $\tilde{A}$, $\tilde{Z}$ and z as in (3.1). Let $r_{\tilde{Z}}$ be defined by

$$r_{\tilde{Z}} = \begin{bmatrix} q_Z \\ - z^T e_{n+1} \end{bmatrix}$$

where q_Z is defined by (5.5). Then $p = \tilde{Z}r_{\tilde{Z}}$ satisfies (6.1), (2.2) and (6.3).

Proof. It is trivial to see that (6.1) holds. We show that (2.2) is satisfied, i.e. p is a descent direction for EMP, as follows:

$$e_{n+1}^T \tilde{Z} \tilde{r}_Z = e_{n+1}^T [Z\ z]\ \tilde{r}_Z$$

$$= e_{n+1}^T Z q_Z - e_{n+1}^T z z^T e_{n+1}$$

$$= - q_Z^T L D L^T q_Z - e_{n+1}^T z z^T e_{n+1} \quad .$$

The first term is zero or negative. The second term is negative since otherwise $\tilde{Z}\tilde{Z}^T e_{n+1} = Z Z^T e_{n+1}$ and hence $\tilde{A}\tilde{\lambda}_L = \hat{A}\lambda_L$, where $\tilde{\lambda}_L$ is λ_L with $(\lambda_L)_j$ deleted, which implies $(\lambda_L)_j = 0$, a contradiction. Thus $\tilde{Z}\tilde{r}_Z$ is a descent direction. The proof that (6.2) holds follows that of Theorem 1, Gill and Murray (1979). We have

$$\hat{a}_j^T \tilde{Z} \tilde{r}_Z = \hat{a}_j^T [Z\ z]\ \tilde{r}_Z = [\ 0 \quad \hat{a}_j^T z]\ \tilde{r}_Z$$

$$= - \hat{a}_j^T z z^T e_{n+1} \ .$$

Note that the fact that $\hat{A}$ has full rank implies that $\hat{a}_j^T \tilde{Z} \neq 0$ and hence $\hat{a}_j^T z \neq 0$. Multiplying both sides of (5.1) by z^T we have

$$(\lambda_L)_j\ z^T \hat{a}_j = z^T e_{n+1}$$

so

$$\hat{a}_j^T \tilde{Z} \tilde{r}_Z = - (\lambda_L)_j\ (\hat{a}_j^T z)^2 > 0 \ . \ \square$$

Note that there is no guarantee that the corresponding range space step $\tilde{Y}\tilde{p}_Y$ to the modified set of constraints satisfies either (2.2) or

(6.2). This step is defined by

$$\tilde{A}^T \tilde{Y} p_{\tilde{Y}} = -\tilde{c}$$

where $\tilde{c}$ is $\hat{c}$ with $\hat{c}_j$ deleted and the orthogonal columns of $\tilde{Y}$ span the range of $\tilde{A}$. However, since $\tilde{Z}r_{\tilde{Z}}$ satisfies (2.2) and (6.2) with strict inequality, so does $\gamma \tilde{Y} p_{\tilde{Y}} + \tilde{Z} r_{\tilde{Z}}$ for small enough γ.

Although $\tilde{Z}r_{\tilde{Z}}$ is guaranteed to satisfy the required properties (6.1), (2.2) and (6.2), the Newton step $\tilde{Z}q_{\tilde{Z}}$ may be preferable if it satisfies (6.2). In fact, the step $\tilde{Z}p_{\tilde{Z}}$, defined by

$$\tilde{Z}^T W \tilde{Z} p_{\tilde{Z}} = - \tilde{Z}^T (e_{n+1} + W \tilde{Y} p_{\tilde{Y}}) ,$$

may be preferable to either, but this is not guaranteed to even be a descent direction. We recommend computing $\tilde{Z}q_{\tilde{Z}}$ and performing the appropriate inner product to check whether it is a feasible descent direction, falling back on $\gamma \tilde{Y} p_{\tilde{Y}} + \tilde{Z} r_{\tilde{Z}}$ if it is not, and substituting $\tilde{Z}q_{\tilde{Z}}$ for $\tilde{Z}r_{\tilde{Z}}$ if it is. This may be done even if $(\mu_W)_j > 0$, which may be the case if we are deleting a constraint when $e_{n+1}^T Y p_Y > 0$, since Theorem 4 does not hold with μ_W substituted for λ_L (even if W is assumed to be positive definite). Although when exact Hessians are available, it is normally inadvisable to delete a constraint and take a Newton step when $(\mu_W)_j > 0$, in this case it is desirable to delete a constraint and the only question is what step to take.

In all of the above, when a constraint is deleted from the active set it is not necessary to recompute the factorizations of $\tilde{A}$ and $\tilde{Z}^T W \tilde{Z}$ from scratch. They can be obtained by updating the factorizations already

available, as described in Gill, Golub, Murray and Saunders (1974) and Gill and Murray (1974).

Note that we have been able to always obtain a satisfactory choice for p by deleting only one constraint with a negative multiplier estimate.

7. Quasi-Newton and Finite Difference Approximations to the Hessian.

It may be that the Hessians $\{\nabla^2 f_i\}$ are unavailable either because they are too expensive to evaluate or too difficult to determine. Here we outline the various alternative ways to approximate W. Recall from Section 1.1 that

$$W = \begin{pmatrix} \bar{W} & 0 \\ 0 & 0 \end{pmatrix}$$

so that when analytical Hessians are used a matrix of order n is stored. The two basic alternatives are using a finite difference approximation or a quasi-Newton approximation.

A finite difference approximation to $\bar{W}$, unlike a quasi-Newton approximation, requires extra gradient evaluations to be done at each iteration. However, it is important to note that extra gradient evaluations of only the t active functions are required, where t may be significantly less than m, the number of functions which must be evaluated at each iteration. Furthermore, since the matrix $\bar{W}$ is not explicitly required to compute p we can form a direct finite difference approximation to $\overline{WZ}$ by differencing the gradient of the Lagrangian function along the columns of $\bar{Z}$. It is also necessary to difference the gradients along Yp_Y to obtain an approximation to WYp_Y (for computing p_Z and μ_W).

Thus the t active gradients must each be evaluated only $n - t + 2$ times, which may be considerably less than the $n + 1$ required to approximate $\bar{W}$.

When a quasi-Newton method is used, two approaches are possible. One is to maintain an approximation $\bar{B}$ to the full matrix $\bar{W}$. The problem with this approach is that $\bar{W}$ may have negative eigenvalues at the solution and so a hereditary positive definite update may not be appropriate. Nonetheless, Powell (1977) has shown that it is possible to maintain a positive definite approximation and still obtain superlinear local convergence. In the case of linearly constrained optimization, Gill and Murray (1974) have suggested maintaining a quasi-Newton approximation to the projected Hessian $\bar{Z}^T\bar{W}\bar{Z}$, a strategy with two major advantages: the matrix to be stored has lower dimension, and because the projected Hessian is positive definite at the solution a positive definite update is appropriate. Murray and Wright (1978) have suggested and numerically investigated a number of variants of a similar approach for problem NCP. An additional difficulty with nonlinear constraints is that the matrix Z is changing at every iteration as well as the matrix W . We note that when this technique is used the vector p_Z (as opposed to q_Z) and the second order multiplier estimate μ_W are not made available. However, even if the full matrix is approximated these vectors are likely to be unreliable since the quasi-Newton approximation is not reliable in the sense that $\bar{B}\bar{p}$ is rarely a close approximation to $\bar{W}\bar{p}$. When a constraint is deleted from the active set, the approximation may be augmented as follows:

$$\tilde{Z}^T B \tilde{Z} = \begin{bmatrix} Z^T B Z & 0 \\ 0 & 1 \end{bmatrix}$$

The resulting direction $\tilde{Z}q_{\tilde{Z}}$ (substituting $\tilde{Z}^T B \tilde{Z}$ for $\tilde{Z}^T W \tilde{Z}$) is then identical to the direction $\tilde{Z}r_{\tilde{Z}}$ described earlier.

The choice between exact Hessians, finite difference of gradients, and a quasi-Newton method will depend on various circumstances. For further details for both finite difference strategies and quasi-Newton strategies the reader is referred to Murray and Wright (1978), where the same topic is discussed with regard to algorithms for solving problem NCP.

One of the prime applications of the minimax algorithm is to data fitting problems in the ℓ_∞ norm (see Section 11). These problems typically have a large number of functions f_i (observations), but a relatively small number of variables. Furthermore if the functions are not too highly nonlinear the number of active constraints at the solution may be close to $n+1$ (it would be exactly $n+1$ if the problem were linear). Thus it may often be that $t \ll m$ and $n+1-t \ll n$, exactly the situation where the finite difference scheme is most efficient. Since the finite difference scheme will normally produce a better direction of search at each iteration than a quasi-Newton method and also has a higher rate of convergence, it will normally take significantly fewer iterations to reach the solution. Thus the additional work at each iteration may result in a substantial saving overall.

It is worth noting that a feasible point algorithm for problem NCP (such as the barrier trajectory method described in Wright (1976)) may have to evaluate all m functions for each column of Z when using a finite difference scheme. The reason for this is to ascertain that each evaluation point is feasible and hence the t active gradients are well-defined. This consideration does not arise with the minimax problem.

8. Determining the Steplength.

In the previous sections we have shown how, given $x^{(k)}$, we may obtain a satisfactory direction of search $\bar{p}$ which is a descent direction for the minimax function. We now discuss how to determine a steplength α to take along $\bar{p}$. Although a simple steplength algorithm may be all that is required to meet convergence criteria for the overall algorithm, from the point of view of efficiency it is important that the step achieve as large a reduction in the value of the function F_M as possible, given a certain limit on the effort to be expended. Since F_M is not differentiable everywhere a steplength algorithm intended for differentiable functions will not be efficient. We therefore use the steplength algorithm which is described at some length in Murray and Overton (1979). This algorithm is designed specifically for the minimax problem and related problems. It includes a facility for varying the limit on the effort to be expended, producing anything from an algorithm which normally takes only a single function evaluation to one which does an exact linear search.

The steplength algorithm requires an initial guess at α, say α_0, where F_M is to be evaluated first. We set α_0 to either one, or the estimated step to the nearest inactive constraint using the linear approximations at $x^{(k)}$, if this is less than one. Thus

$$\alpha_0 = \min \{1, \alpha_0'\} \tag{8.1}$$

where $\alpha_0' = \min \{ -\frac{c_i}{a_i^T p} \mid a_i^T p < 0 \text{ and } i \text{ not in active set} \}$.

It is possible that $\alpha_0 = 0$ and hence the algorithm must allow for this possibility. In this case the appropriate inactive constraint must be added to the active set and the iteration is repeated.

9. Flowchart of the Algorithm.

We summarize the basic iteration of the algorithm in the flowchart in Figure 2. For simplicity we have omitted any tolerances from the flowchart though clearly in practice these must be included. The parameters γ_1 and γ_2 are optional. For the results presented in Section 14, γ_1 is set to 1 when possible but γ_2 is always set to zero. The latter is done because near a saddle point or after deleting a constraint a poor range space direction can swamp a good null space direction if $\gamma_2 > 0$. The notation S is used to mean either W or a finite-difference or quasi-Newton approximation to W.

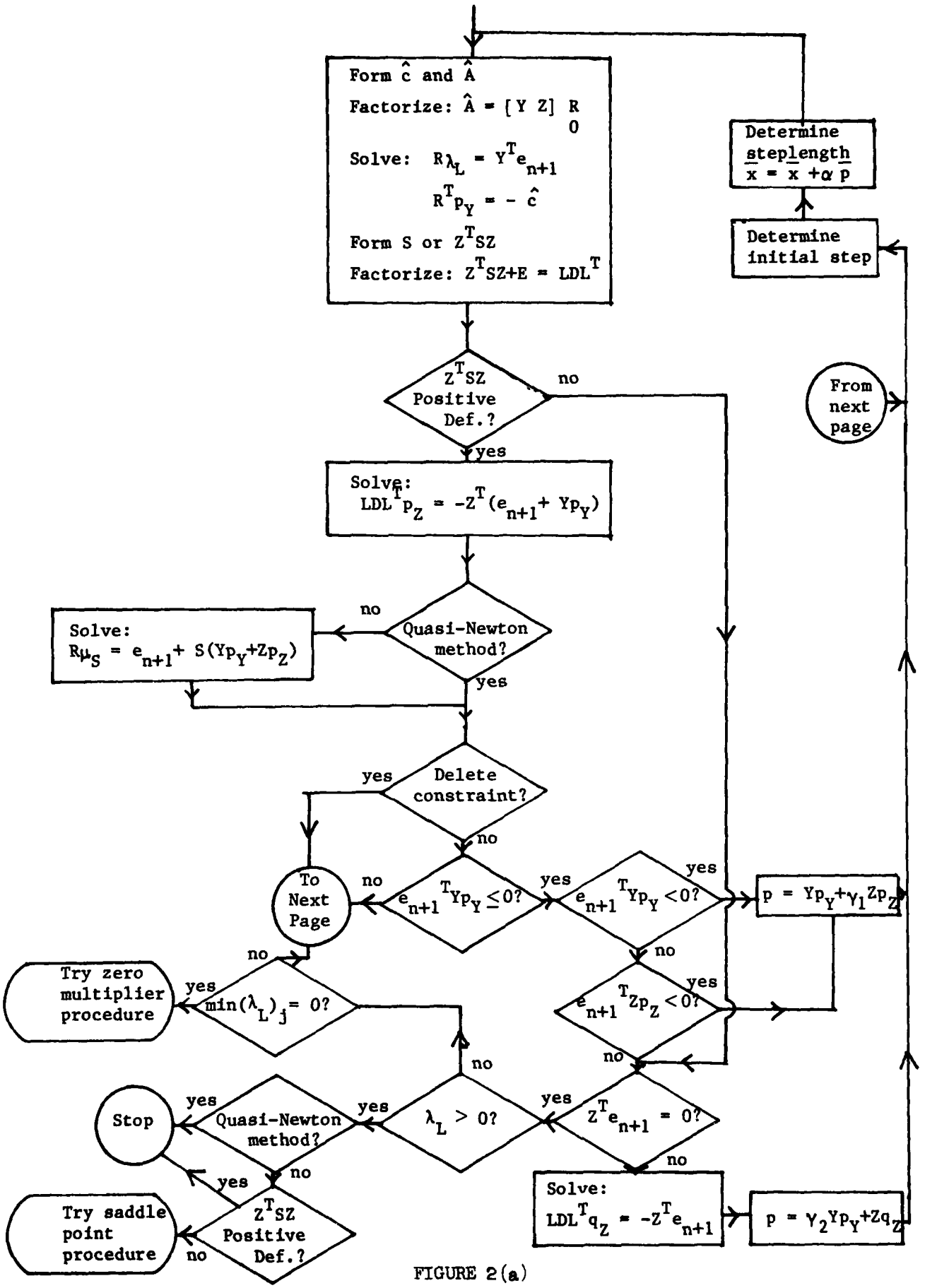

Form $\hat{c}$ and $\hat{A}$
Factorize: $\hat{A} = [Y\ Z]\begin{bmatrix} R \\ 0 \end{bmatrix}$
Solve: $R\lambda_L = Y^T e_{n+1}$
$R^T p_Y = -\hat{c}$
Form S or Z^TSZ
Factorize: $Z^TSZ+E = LDL^T$
Z^TSZ Positive Def.?
no
yes
Solve: $LDL^T p_Z = -Z^T(e_{n+1} + Yp_Y)$
Quasi-Newton method?
no
Solve: $R\mu_S = e_{n+1} + S(Yp_Y+Zp_Z)$
yes
Delete constraint?
yes
no
To Next Page
$e_{n+1}^T Yp_Y \leq 0$?
no
yes
$e_{n+1}^T Yp_Y < 0$?
yes
$p = Yp_Y + \gamma_1 Zp_Z$
no
$e_{n+1}^T Zp_Z < 0$?
yes
no
$\min(\lambda_L)_j = 0$?
yes
Try zero multiplier procedure
no
$\lambda_L > 0$?
yes
no
$Z^T e_{n+1} = 0$?
yes
no
Solve: $LDL^T q_Z = -Z^T e_{n+1}$
$p = \gamma_2 Yp_Y + Zq_Z$
Quasi-Newton method?
yes
Stop
no
Z^TSZ Positive Def.?
yes
no
Try saddle point procedure
Determine steplength $x = x + \alpha p$
Determine initial step
From next page

FIGURE 2(a)

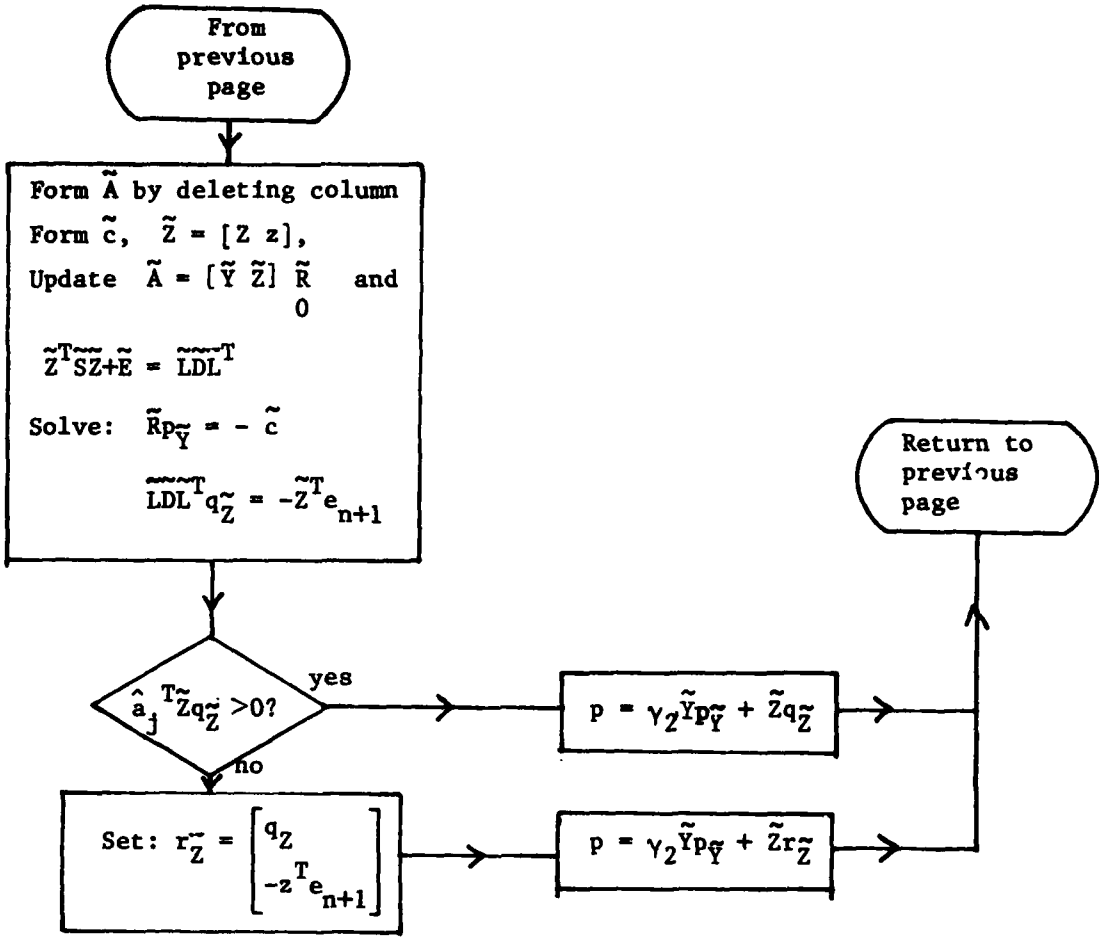

From previous page
Form $\tilde{A}$ by deleting column
Form $\tilde{c}$, $\tilde{Z} = [Z\ z]$,
Update $\tilde{A} = [\tilde{Y}\ \tilde{Z}] \begin{matrix} \tilde{R} \\ 0 \end{matrix}$ and
$\tilde{Z}^T S \tilde{Z} + \tilde{E} = \tilde{L}\tilde{D}\tilde{L}^T$
Solve: $\tilde{R} p_{\tilde{Y}} = -\tilde{c}$
$\tilde{L}\tilde{D}\tilde{L}^T q_{\tilde{Z}} = -\tilde{Z}^T e_{n+1}$
$\hat{a}_j^T \tilde{Z} q_{\tilde{Z}} > 0$?
yes
no
$p = \gamma_2 \tilde{Y} p_{\tilde{Y}} + \tilde{Z} q_{\tilde{Z}}$
Set: $r_{\tilde{Z}} = \begin{bmatrix} q_Z \\ -z^T e_{n+1} \end{bmatrix}$
$p = \gamma_2 \tilde{Y} p_{\tilde{Y}} + \tilde{Z} r_{\tilde{Z}}$
Return to previous page

FIGURE 2(b)

10. Selecting the Active Set.

The success of the algorithm we have described depends on being able to make a reasonable choice of the active set at each iteration. If constraints are required to have very small magnitude to be included in the active set, then the iterates will follow the constraint boundaries very closely and the convergence will be slow. Conversely if too many constraints are selected as active the directions of search may be poor. In this section we describe an active set strategy that has been most successful for the numerical experiments we have carried out. It is likely, however, that this strategy will be modified in the future after more extensive numerical experience.

Clearly one feature required of the active set strategy is that, as the iterates approach $\overset{*}{x}$, it should become successively more difficult for constraints with magnitudes significantly greater than zero to be included. One way to accomplish this is to have the strategy depend on a parameter which is reduced as the solution is approached or if any difficulties arise. We found that reducing a parameter in this way was not satisfactory since it can easily happen that the parameter is reduced too much, making it impossible to ever include all the correct active constraints. Instead we take the following approach. There is always at least one constraint active with value identically zero so the first decision is whether to include a second constraint. If this is done the decision of whether to include others is tied to the magnitude of the second constraint. Thus the required objective will be achieved, provided that the first decision is made correctly so that any second active

constraint approaches zero as the iterates approach the solution.

Let us order the constraints so that $0 = c_1 \leq c_2 \leq \ldots \leq c_m$, with the $\{f_i\}$ and $\{v_i\}$ correspondingly ordered. We include in the active set all constraints with magnitude less than a very small tolerance, any κ_0. In order to compare the larger constraint magnitudes we must scale them in some way. For problem $\ell_\infty P$ (see Section 11) we define the scaled values by $\bar{c}_i = \frac{c_i}{F_\infty}$, since the value of F_∞ can only be very small if all the constraints values are very small. For problem MMP the value of F_M could be zero or negative so we define the scaled values by $\bar{c}_i = \frac{c_i}{1+|f_1|+|f_2|}$. Scaling the values is necessary since two functions with values 999 and 1000 are likely to both be active if one is, while this is not true of functions with values 1 and 2. Notice that the existence of any functions in problem MMP with negative values much less than F_M does not affect the definition of $\bar{c}_i$, which is appropriate since it does not affect the solution.

If only one constraint is active, then $\|v_1(\bar{x}^*)\| = 0$. Therefore the decision of whether to include a second active constraint c_2 is made as follows. If $\|v_1\|$ is small, say $\|v_1\| < \kappa_1$, then c_2 is included only if $c_2 < \kappa_2\|v_1\|^2$, where $\kappa_2 > 1$, a test which can be justified by Taylor expansions around the solution. If $\|v_1\| \geq \kappa_1$, then c_2 is included only if $\bar{c}_2 < \kappa_3$, where $0 < \kappa_3 < 1$.

It remains to test the other constraint values against c_2. We include c_i if c_{i-1} is included and $\bar{c}_i < \kappa_3$ and $\bar{c}_i < (\bar{c}_{i-1})^{\kappa_4}$, $i = 3,4,\ldots, \min(m,n+1)$, where $0 < \kappa_4 < 1$. Thus, for example, if

$\kappa_2 = 0.5$, c_3 is included if c_2 is included, $\bar{c}_2 = 10^{-6}$ and $\bar{c}_3 = 10^{-4}$, but not if $\bar{c}_2 = 0.1$ and $\bar{c}_3 = 0.5$. Note that we always have $\bar{c}_2 \leq 1$.

We found that it was also important not to include a constraint in the active set if it was included but subsequently deleted in the previous iteration. It was also helpful to include a constraint requiring only that $\bar{c}_i < \kappa_3$ if the previous line search chose a step to a point of discontinuity involving the corresponding function.

Sometimes some constraints are much smaller than others which should also be in the active set (for example if some of the functions are linear). In such a situation it is not appropriate to compare all the constraints with c_2 . The remedy is to let the first constraint with value significantly greater than the machine precision play the role of c_2 . However, then v_1 must be replaced by a projected gradient $Z^T v_1$ in the tests, and the factorization of $\hat{A}$ and the accumulation of the transformations which form Z must be performed as the active set selection proceeds. We have not yet implemented this modification.

Our current active set strategy is the above with κ_0 set to the square root of the machine precision, $\kappa_1 = 0.1$, $\kappa_2 = 25$, $\kappa_3 = 0.25$ for $\ell_\infty P$, $\kappa_3 = 0.1$ for MMP, and $\kappa_4 = 0.5$.

11. The Special Case of the Least ℓ_∞ Norm Problem.

In this section we consider problem $\ell_\infty P$, defined in Section 1. Problem $\ell_\infty P$ is an important special case of MMP. Note that $\ell_\infty P$ must have a solution, whereas MMP may not. Clearly it is preferable to treat $\ell_\infty P$ in a special manner rather than just treat $|f_i|$ as $\max(-f_i, f_i)$. If we eliminate the possibility that F_∞ is zero at the solution, i.e. <u>all</u> the $\{f_i\}$ are zero at the solution, then we can observe that only <u>one</u> of each pair of constraints $x_{n+1} - f_i(\bar{x}) \geq 0$, $x_{n+1} + f_i(\bar{x}) \geq 0$, can be active at the solution. Thus defining $\sigma_i = \text{sgn}(f_i(\bar{x}))$ for any $\bar{x}$ it is straightforward to handle $\ell_\infty P$ by using the algorithm described for MMP, replacing f_i by $\sigma_i f_i$ everywhere. The only places where it is necessary to consider the <u>inactive</u> constraints $\{c_i' = x_{n+1} + \sigma_i f_i\}$ are in the determination of the initial guess at the steplength α_0, and in the steplength algorithm itself. Thus α_0 must be set to

$$\alpha_0 = \min\{1, \alpha_0', \alpha_0''\}$$

where α_0' is defined by (8.1) and α_0'' is given by

$$\alpha_0'' = \min\left\{-\frac{c_i'}{\nabla c_i'^T p} \;\middle|\; \nabla c_i'^T p < 0, \quad i = 1,2,\ldots,m\right\}.$$

The changes which must be made to the steplength algorithm are indicated in Murray and Overton (1979).

If F_∞ is zero at the solution there is still no difficulty with this approach provided that $m \geq n+1$, since then of the 2m constraints active for the equivalent problem corresponding to $\ell_\infty P$, at most $n+1$ can be

included in the active set to obtain a full rank active constraint Jacobian. Thus including in the active set at most one constraint of each pair causes no difficulty. Such a situation is a highly degenerate one but the point is that if the situation does arise the algorithm will take care of it efficiently. The usual source of $\ell_\infty P$ is data fitting problems so we expect almost always to have $m \geq n+1$. However, if it does happen that we wish to solve $\ell_\infty P$ with $m \leq n$ then the above technique may be very inefficient since both constraints in a pair of constraints active at the solution cannot be put into the active set. Instead of making a complicated modification to take care of this unlikely possibility, we recommend writing the problem explicitly in the form MMP and solving this directly.

It is straightforward to generalize the above to an algorithm which can be used to minimize a more general function

$$F_{GM}(\bar{x}) = \max \{ \max_{1 \leq i \leq m_1} |f_i(\bar{x})|, \max_{m_1+1 \leq i \leq m} f_i(\bar{x}) \},$$

assuming $m_1 \geq n+1$ if $m_1 \neq 0$. Since coping with the general case introduces very little extra overhead our implementation of the algorithm handles this wider class of functions. In this way one algorithm takes care of both MMP ($m_1 = 0$) and $\ell_\infty P$ ($m_1 = m$).

11.1 The Haar Condition.

We comment here on the meaning of the Haar condition since this is usually discussed in the context of ℓ_∞ approximation. Let us first consider the case that the f_i are linear. The Haar condition is said

to hold for these functions if every $n \times n$ submatrix of V is non-singular, where V is the $n \times m$ matrix whose columns are the gradients of the $\{f_i\}$. Consider problem EMP which is now a linear programming problem. A necessary condition for $\bar{x}^*$ to be a solution to EMP is that $\hat{V}\hat{\lambda}^* = 0$ and $\hat{\lambda}^* \neq 0$ (since $\hat{e}^T \hat{\lambda}^* = 1$). Thus the requirement that the Haar condition holds implies that there be at least $n+1$ active constraints at the solution with none of the multipliers equal to zero, and hence that the solution is unique. Most algorithms which solve the linear minimax problem do so by solving a linear program related to ELP. Thus whether or not the Haar condition holds is quite irrelevant to the difficulty of solving $\ell_\infty P$, since zero multipliers cause no real difficulty in solving linear programs. Degeneracy, which occurs when the matrix $\hat{A}$ is rank deficient and can in theory cause problems in solving a linear program, can occur whether or not the Haar condition holds. The significance of the Haar condition is that if it does not hold the solution may not be unique, and hence one may be interested in a "strict" solution to the data approximation problem which is unique, i.e. that solution of $\ell_\infty P$ which minimizes

$$\max \{|f_i(\bar{x})| \;\Big|\; f_i \text{ inactive at } \bar{x}^*\}$$

(see Rice (1969) and Brannigan (1978)).

The Haar condition is much stronger than necessary to ensure uniqueness. A slightly more reasonable condition is that there be $n+1$ constraints with nonzero multipliers active at the solution of $\ell_\infty P$ (see Jittorntrum and Osborne (1979, p.3) for an example showing that this condition is

still stronger than necessary to ensure uniqueness). This condition cannot be checked without solving $\ell_\infty P$ but solving $\ell_\infty P$ is actually much easier to do than checking whether the Haar conditon holds.

Let us now consider the nonlinear case. The Haar condition is then said to hold at a point $\bar{x}$ if every subset of gradients of functions with zero constraint values at $\bar{x}$ is linearly independent. As in the linear case we are really only concerned with the situation at $\overset{*}{\bar{x}}$. Thus the Haar condition holds at $\overset{*}{\bar{x}}$ if every $n \times n$ submatrix of $\overset{*}{V}$ has full rank. Again it follows that if the Haar conditon holds there must be $n+1$ constraints with nonzero multipliers active at $\overset{*}{\bar{x}}$. This condition is however a much more unreasonable one than in the linear case. There must always exist a solution to the linear problem where $n+1$ constraints are active, but the nonlinear problem can have a unique solution with anything from 1 to $n+1$ constraints active. If the Haar condition does hold at $\overset{*}{\bar{x}}$ then $\overset{*}{Z}$ is null and the problem can be adequately solved by a method using only first-order information. Thus our algorithm has been designed with the assumption that the Haar condition often does not hold. When $\ell_\infty P$ arises from data fitting problems it may sometimes be the case that the Haar condition can be expected to hold at $\overset{*}{\bar{x}}$. However, the only way to ascertain whether the Haar condition does indeed hold is to solve the problem. Since we cannot be sure at the outset it is clearly unsatisfactory to be using an algorithm which uses first order-information only and hence may indicate that the Haar condition does not hold by extremely poor performance. For many data fitting problems the number of active constraints at the solution may be <u>close</u> to $n+1$, and as pointed

out in Section 7 this means using a finite difference approximation to Z^TWZ, which has order $n+1-t$, may be quite inexpensive.

Cromme (1972) discusses in a broader setting a weaker condition than the Haar condition at $\bar{x}^*$ called strong uniqueness which still ensures that a particular algorithm using only first-order information converges quadratically. Strong uniqueness implies that $n+1$ constraints are active at $\bar{x}^*$ but is weaker than the Haar condition since it permits active constraints with zero multipliers. Jittorntrum and Osborne (1979) discuss a still weaker condition which arises from examining the curvature in the null space of the active constraint Jacobian with gradients corresponding to zero multipliers deleted. This weaker condition also ensures that a first-order method has quadratic convergence. The sufficient conditions for a minimum given in Section 1.1 are much weaker than any of these conditions since they permit $t(\bar{x}^*)$ to be less than $n+1$, in which case an efficient algorithm must approximate Z^TWZ. These conditions could themselves be weakened if curvature corresponding to zero multipliers were examined.

12. Relationships to Other Algorithms

A number of other algorithms have been proposed to solve the nonlinear minimax problem. It was not until recently, however, that special algorithms for MMP which make use of second-order information appeared. Han (1977b, 1978a, 1978b) suggests methods which solve a sequence of quadratic programming problems using a quasi-Newton approximation to $\bar{W}$. These will discussed further shortly. Watson (1979) and Hald and Madsen (1978) proposed two-stage methods which begin by using the first-order methods of Anderson and Osborne (1977) and Madsen (1975), respectively, and switch to solving a system of nonlinear equations using the second-order information in $\bar{W}$ when it is thought that the active set has been identified. The system is of order $n+1+t$, i.e., the multipliers and variables are all obtained at once. Recall that for problem $\ell_\infty P$, t is often close to $n+1$ so the systems that Watson and Hald and Madsen solve may be much larger than the ones we solve. The direction of search obtained is not necessarily a descent direction for the minimax function but only for the residual of the nonlinear system. A method related to the second stage of these methods was given by Hettich (1976). Conn (1979) presents a method which is derived from the point of view of a nondifferentiable penalty function. It is related to the algorithm of Charalambous and Conn (1978) but uses second-order information. We discuss this method further below. Other methods which use second-order information and are related to Han's method are disucssed by Charalambous and Moharram (1978,1979) and Wierzbicki (1978).

Our algorithm is most closely related to the methods of Han (1977b, 1978a) and Conn (1979), so we discuss these further here. The primary difference between our method and Han's method is that we attempt to identify the active set at each iteration and then solve an equality-constrained quadratic program (EQP), modifying the resulting direction of search and the active set if necessary, while Han solves an inequality-constrained quadratic program (IQP), thus implicitly selecting the active set associated with the solution of IQP. The IQP has the form:

$$\min \frac{1}{2} p^T W p + e_{n+1}^T p$$

$$\text{subject to } A^T p \geq -c$$

where c and A are the vector and matrix of all the constraints and their gradients.

The dichotomy of whether to solve EQP or IQP occurs at all levels of constrained optimization. Murray (1969a) considered solving the IQP associated with his algorithm for NCP but found an EQP strategy more successful. He also considered a strategy of partially solving IQP. The same question arises in linearly constrained optimization. Brayton and Cullum (1977) report some results which indicate that for the case of minimization subject to bounds on the variables (the simplest possible constrained optimization problem), solving IQP is not in general more efficient than solving EQP.

The motivation for solving IQP is that it makes the fullest use of the information at $x^{(k)}$. Furthermore it simplifies the description of the algorithm and for problem MMP makes it straightforward to get a descent direction since positive constraints are not selected for an active set except in the process of solving IQP. Clearly one disadvantage of solving IQP is that it is more work than solving EQP. If $m \gg n$ and the function evaluations are not too expensive then an algorithm which solves IQP may be extremely inefficient compared to one which solves EQP. However this is not the main objection to solving IQP. The main motivation for solving EQP rather than IQP is that the linear approximations to the constraints (for NCP and MMP) and the quadratic approximation to the Lagrangian function are unreliable away from the current point $x^{(k)}$. The process of solving IQP involves successively making decisions about which constraints to include in the active set at points which may be quite far from $x^{(k)}$, based on the approximations at $x^{(k)}$. Thus the final point at which this decision is made, the solution of IQP, may be the result of choosing an active set which has no meaning whatsoever. If $x^{(k)}$ is so close to x^* that the approximations are satisfactory then IQP may still have no advantage over EQP since they may well have the same solution.

Most of the differences between Han's method and ours result from the difference between IQP and EQP (he uses the multipliers from the old IQP and requires that the full Hessian $\bar{W}$ is positive definite, while we use λ_C to define $\bar{W}$ and require only that Z^TWZ be positive definite). He discusses only quasi-Newton methods while we consider also a finite difference strategy. Finally, although his line search is used to obtain a reduction in F_M it is not designed specially for nondifferentiable functions as ours is.

Han (1978b) presents another algorithm for MMP which is related to the one discussed above. It is quite different however in that the line search takes place in the (n+1) dimensional space. The line search obtains a reduction in the function

$$\theta(\alpha) = x_{n+1} + \alpha p_{n+1} + \sum_{i=1}^{m} \max(f_i(\bar{x} + \alpha \bar{p}) - (x_{n+1} + \alpha p_{n+1}), 0)$$

instead of $F_M(\bar{x} + \alpha \bar{p})$. The motivation given for this is that θ takes into account some inactive functions while F_M gives bias completely to the active functions. However, our view is that inactive functions should be considered in the line search only if they are likely to become active along the line, and this is exactly what our special line search to reduce the minimax function does.

The problem with reducing θ is that it relies too much on the value of p_{n+1} which gives only a linear approximation to the active functions at $x^{(k)}$. It is easy to construct examples where minimizing θ along the line results in a much smaller reduction of F_M than is possible. We feel that using θ instead of F_M in the line search is discarding one of the most useful tools available to solve MMP.

The analysis in Han (1978b) is concerned with the fact that the quadratic form of IQP is not positive definite in R^{n+1}. Our view however is that the only matrix whose positive definiteness should be a concern is the projected Hessian, and this is the same in R^n and R^{n+1} since $\bar{Z}^T \bar{W} \bar{Z} = Z^T W Z$. Recall that we expect the dimension of this matrix to be much smaller than that of W.

We now turn our attention to the algorithm of Conn (1979), which is more closely related to ours in a number of ways. Like us, Conn attempts to identify the active set and solves a related EQP at every iteration. However, unlike ours, his search direction does not include a component in the space spanned by Y unless there is reason to believe that $x^{(k)}$ is near a stationary point. Furthermore he uses the Hessian of the Lagrangian function to give the quadratic form of the QP only if $x^{(k)}$ is thought to be near a stationary point. Otherwise he uses instead the Hessians of <u>one</u> of the active functions at each iteration. His reason for this is that the Lagrange multipliers may be highly inaccurate away from a stationary point. Although this is certainly true, our view is that using the Hessian of only one function is equivalent to using a multiplier estimate with only one component equal to one and the rest zero, and that this is no less arbitrary than using any other vector of nonnegative components which sum to one to define W. As we explained at the end of Section 4, our algorithm always uses such a vector to define W. Our approach eliminates any need to decide when to switch from one strategy to another, something which it is difficult to do since it is hard to tell how accurate multiplier estimates are. Furthermore using different Hessians at different iterations makes a quasi-Newton approach difficult.

There are many other differences between Conn's algorithm and ours which follow because of the fact that his approach is via a nondifferentiable penalty function while ours is via a Langrangian function. For example, he does not factorize the matrix $\hat{A}$ as we do, but instead factorizes a matrix M which is $-\hat{V}$ less one column $\hat{v}_j$ corresponding to the one function whose Hessian is to be computed. This matrix factorization

is then updated to give a factorization of the matrix resulting from adding $\hat{v}_j$ to each column of M. It can be shown that this approach restricts $\bar{p}$ to the same null space as our algorithm. Multiplier estimates can also be computed by this approach but they will not be the same as either λ_L or λ_C since they give bias to function j. Conn shows how to take advantage of any explicitly linear functions in his algorithm.

In a way our algorithm treads the middle ground between Han's method and Conn's method. Han relies on the approximation at $x^{(k)}$ so completely that he solves the IQP. Conn distrusts the multiplier estimates and does not use them unless $x^{(k)}$ is near a stationary point. We believe that some multiplier estimates are better than no estimates at all, but we solve the QP which relies as little on the approximations as possible.

13. Convergence Properties.

In the limit our method becomes a projected Lagrangian method for NCP applied to the special case MMP and so we can make use of the known asymptotic local convergence results for these methods. Robinson (1974) showed that the method of Wilson (1963) has a quadratic rate of local convergence if analytical Hessians of the objective function and constraints are used, provided that the functions are sufficiently smooth and there are no zero multipliers at the solution. This last condition also ensures that ultimately solving the IQP and the EQP are the same, so the only difference between our method and Wilson's in the limit is the fact that we use the first-order multiplier estimates at the new point (instead of the multipliers of the old QP) to define W . It is shown by Fletcher (1974) that using the first-order multiplier estimates does not inhibit the quadratic convergence of a projected Lagrangian method.

Using finite-difference approximations to the Hessian or projected Hessian is well known to have essentially the same final rate of convergence (to machine precision) as using analytical Hessians. A superlinear rate of convergence for a projected Lagrangian method using a particular quasi-Newton approximation to the full matrix W has been shown by Powell (1977).

14. Computational Results.

In this section we illustrate the usefulness of the algorithm by presenting some numerical results for the case where finite-difference approximations to the Hessian of the Lagrangian function are used.

Six problems are selected from the literature. We solve the first four as problems of the type $\ell_\infty P$ and the last two as problems of the type MMP. The problems and their solutions are as follows:

Problem 1: Bard (1970).

$$f_j(x) = y_j - x_1 - \frac{u_j}{v_j x_2 + w_j x_3} \quad , \quad j = 1,2,\ldots,15$$

where $u_j = j$, $v_j = 16-j$, $w_j = \min(u_j, v_j)$ and the vector y is given in Table 1.

Starting Point: $x_0 = (1,1,1)^T$.

Problem 2: Kowalik and Osborne (1968).

$$f_j(x) = y_j - \frac{x_1(u_j^2 + x_2 u_j)}{u_j^2 + x_3 u_j + x_4} \quad , \quad j = 1,2,\ldots,11$$

where vectors y and u are given in Table 2.

Starting Point: $x_0 = (0.25,0.39,0.415,0.39)^T$

TABLE 1

i	y_i
1	0.14
2	0.18
3	0.22
4	0.25
5	0.29
6	0.32
7	0.35
8	0.39
9	0.37
10	0.58
11	0.73
12	0.96
13	1.34
14	2.10
15	4.39

TABLE 2

i	y_i	u_i
1	0.1957	4.0000
2	0.1947	2.0000
3	0.1735	1.0000
4	0.1600	0.5000
5	0.0844	0.2500
6	0.0627	0.1670
7	0.0456	0.1250
8	0.0342	0.1000
9	0.0323	0.0833
10	0.0235	0.0714
11	0.0246	0.0625

Problem 3: Madsen (1975).

$$f_1(x) = x_1^2 + x_2^2 + x_1 x_2$$

$$f_2(x) = \sin x_1$$

$$f_3(x) = \cos x_2 \quad .$$

Starting Point: $x_0 = (3,1)^T$.

Problem 4: El-Attar et al. (1979) # 2.

$$f_1(x) = x_1^2 + x_2^2 + x_3^2 - 1 \qquad f_4(x) = x_1 + x_2 - x_3 + 1$$

$$f_2(x) = x_1^2 + x_2^2 + (x_3 - 2)^2 \qquad f_5(x) = 2x_1^3 + 6x_2^2 + 2(5x_3 - x_1 + 1)^2$$

$$f_3(x) = x_1 + x_2 + x_3 - 1 \qquad f_6(x) = x_1^2 - 9x_3$$

Starting Point: $x_0 = (1,1,1)^T$.

Problem 5: Charalambous and Bandler (1976) # 1.

$$f_1(x) = x_1^2 + x_2^4$$

$$f_2(x) = (2-x_1)^2 + (2-x_2)^2$$

$$f_3(x) = 2 \exp(-x_1 + x_2) \quad .$$

Starting Point: $x_0 = (1,-0.1)^T$.

Problem 6: Rosen and Suzuki (1965).

$$f_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 .$$

$$f_2(x) = f_1(x) - 10(- x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 + 8)$$

$$f_3(x) = f_1(x) - 10(- x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 + 10)$$

$$f_4(x) = f_1(x) - 10(- 2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 + 5) .$$

Starting Point: $\bar{x}_0 = (0,0,0,0)^T$.

(This is a problem originally of the type NCP transformed to type MMP by the introduction of the penalty parameter 10, which is always possible if the parameter is large enough, as several authors have pointed out).

Solutions found:

Problem 1, (ℓ_∞P):

$$F_\infty(\overset{*}{x}) = 0.77601 \quad \text{with} \quad \overset{*}{x} = (0.17757, - 0.94295, 5.30796)^T .$$

(This is a different local minimum from that found by Watson (1979)).

Problem 2, (ℓ_∞P):

$$F_\infty(\overset{*}{x}) = 0.0080844 \text{ with } \overset{*}{x} = (0.18463, 0.10521, 0.01196, 0.11179)^T.$$

Problem 3, (ℓ_∞P):

$$F_\infty(\overset{*}{x}) = 0.61643 \quad \text{with} \quad \overset{*}{x} = (- 0.45330, 0.90659)^T .$$

Problem 4, (ℓ_2P):

$$F_2(\overset{*}{x}) = 3.59972 \quad \text{with} \quad \overset{*}{x} = (0.32826,\ 0.00000,\ 0.13132)^T$$

Problem 5, (MMP):

$$F_M(\overset{*}{x}) = 1.95222 \quad \text{with} \quad \overset{*}{x} = (1.13904,\ 0.89956)^T$$

Problem 6, (MMP):

$$F_M(\overset{*}{x}) = -44.0000 \quad \text{with} \quad \overset{*}{x} = (0.00000,\ 1.00000,\ 2.00000,\ -1.00000).$$

The results are summarized in Table 3. The termination conditions requires were that $\|c\|_2 < 10^{-6}$, $\|Z^T e_{n+1}\|_2 < 10^{-6}$, $Z^T W Z$ numerically positive semi-definite and $\lambda_C \geq 0$. The line search accuracy parameter η was set to 0.9 (see Murray and Overton (1979) for the definition of this parameter). Several other choices of η were tried, but $\eta = 0.9$ was the most efficient, indicating as expected that a slack line search is desirable at least on these problems. The machine used was an IBM 370/168 in double precision, i.e., with 16 decimal digits of accuracy. The column headed NI reports the number of iterations required, which is also the number of times the Hessian was approximated. The column headed NF gives the number of function evaluations (not including gradient evaluations for the Hessian approximation).

TABLE 3

Problem	n	m	$n+1-t^*$	NI	NF
1 (Bard)	3	15	1	11	11
2 (Kowalik and Osborne)	4	11	0	11	14
3 (Madsen)	2	3	1	13	19
4 (El-Attar et al,#2)	3	6	2	7	8
5 (Charalambous and Bandler,#1)	2	3	1	6	6
6 (Rosen and Suzuki)	4	4	2	7	10

The results demonstrate that at least on a limited set of test problems the algorithm fulfills some of its promise. Final quadratic convergence was observed in all cases. The algorithm has been tested on a wider set of problems and results obtained for a variety of choices of the optional parameters. It was clear from these more extensive results that more work needs to be done in developing the active set selection strategy. These results must therefore be regarded as preliminary.

15. Concluding Remarks.

We conclude with emphasizing the importance of solving MMP by a special algorithm such as the one presented here and not just applying an algorithm for the general nonlinear programming problem NCP to the equivalent problem EMP. The primary simplification of the minimax problem over the nonlinear programming problem is that a natural merit function is available to measure progress towards the solution. To put this another way, it is always possible to reduce F_M in the line search and

obtain a new feasible point for problem EMP. Several results in this chapter followed from the availability of the natural merit function. In particular, consider Theorem 4 (Section 5.1). This shows that if YP_Y, the component of the search direction in the range space of the active constraint Jacobian, is an uphill direction with respect to the minimax function, then it is known that too many constraints are in the active set and a constraint with a positive value and a negative multiplier estimate can be deleted to obtain a descent direction. There is no analogue of Theorem 4 for the nonlinear programming problem NCP, because $\hat{c}$ may have negative components. If the vector YP_Y is uphill with respect to an artificial merit function such as a penalty function, then it may be because there are too many constraints in the active set, or it may be because the penalty parameter is not large enough.

There are other aspects which make it clear that solving MMP in a special way is advantageous. Since the first-order constraint qualifications are always satisfied (see Section 1.2) there is no need to be concerned over the existence of Lagrange multipliers when the active constraint Jacobian becomes rank deficient. Also, as was pointed out in Section 4, problem MMP is in some sense naturally scaled and the first-order Lagrange multiplier estimates can take advantage of this fact.

It should be clear by now how our algorithm is related to the projected Lagrangian algorithms which have been proposed to solve NCP. Wilson (1963), Robinson (1974), Han (1977a) and Powell (1977) all solve successive inequality constrained QP's, so in that sense they are more closely related to the method of Han (1977b, 1978a) than to our method. Murray (1969a, 1969b), Wright (1976) and Murray and Wright (1978) solve

successive equality-constrained QP's. However, their methods differ from the others and from ours in the sense that they do not attempt to step to the active constraint boundaries at every step but control how far outside or inside the feasible region the iterates stay by means of penalty and barrier parameters. This type of approach has proved to be very successful for solving NCP because it balances the reduction of the objective function with the reduction of the constraint violation in a satisfactory way. However, this approach is quite unnecessary for solving MMP since it is always trivial to obtain a feasible point for EMP. To put it another way, reducing the minimax function in the line search always results in a step towards the constraint boundaries, although we do not usually wish to step exactly to the boundaries by doing an exact line search.

Constrained Problems.

Linear constraints can be handled by the algorithm we have presented, since they can be incorporated into the QP at each iteration. It follows from the above remarks however that nonlinear constraints cannot be handled by the algorithm for MMP in a straightforward way. As soon as nonlinear constraints are introduced the natural merit function is lost and the problem takes on the complexity of the general nonlinear programming problem NCP. Of course nonlinear constraints can still be handled by nonlinear programming methods, but it is important to recognize the increase in complexity. Clearly the best approach would be one which takes advantage of the minimax structure and introduces an artificial merit function dealing with the genuine nonlinear constraints and not with those of EMP.

REFERENCES.

Anderson, D.H. and M.R. Osborne (1977). Discrete, nonlinear approximations in polyhedral norms: a Levenberg-like algorithm, Num. Math. 28, pp. 157-170.

Bard, Y. (1970). Comparison of gradient methods for the solution of nonlinear parameter estimation problems, SIAM J. Num. Anal. 7, pp. 157-186.

Brannigan, M. (1978). The strict Chebyshev solution of overdetermined systems of linear equations with rank deficient matrix, National Research Institute for Mathematical Sciences of the CSIR, Report, Pretoria, Rep. South Africa.

Brayton, R.K. and J. Cullum (1977). Optimization with the parameters constrained to a box, Proc. IMACS Intl. Symp., Simulation Software and Numerical Methods for Stiff Differential Equations.

Charalambous, C. and J.W. Bandler (1974). Nonlinear minimax optimization as a sequence of least p-th optimization with finite values of p, Internat. J. Systems Sci. 7, pp. 377-391.

Charalambous, C. and A.R. Conn (1978). An efficient method to solve the minimax problem directly, SIAM J. Numer. Anal. 15, pp. 162-187.

Charalambous, C. and O. Moharram (1978). A new approach to minimax optimization, in G.J. Savage and P.H. roe, eds., Large Eng. Systems 2, pp. 169-172.

Charalambous, C. and O. Moharram (1979). Quasi-Newton methods for minimax optimization, Dept. of Systems Design Report 48-0-240179, University of Waterloo.

Cheney, E.W. (1966). Introduction to Approximation Theory, McGraw-Hill, New York.

Conn, A.R. (1979). An efficient second-order method to solve the (constrained) minimax problem, Dept. of Combinatories and Optimization Rept. CORR-79-5, University of Waterloo.

Cromme, L. (1978). Strong uniqueness: A far reaching criterion for the convergence analysis of iterative procedures, Num. Math. 29, pp. 179-194.

Demyanov, V.F. and V.N. Malozemov (1974). Introduction to Minimax, Wiley, New York and Toronto.

El-Attar, R.A., M. Vidyasagar and S.R.K. Dutta (1979). An algorithm for L_1 norm minimization with application to nonlinear L_1 approximation, SIAM J. Num. Anal. 16, pp. 70-86.

Fiacco, A.V. and G.P. McCormick (1968). Nonlinear Programming: Sequential Unconstrained Optimization Techniques, Wiley, New York.

Fletcher, R. (1974). Methods related to Lagrangian functions, in P.E. Gill and W. Murray, eds., Numerical Methods for Constrained Optimization, Academic Press, New York and London, pp. 219-240.

Fletcher, R. and T.L. Freeman (1977). A modified Newton method for minimization, JOTA 23, pp. 57-372.

Gill, P.E., G.H. Golub, W. Murray, and M.A. Saunders (1974). Methods for modifying matrix factorizations, Math. Comp. 28, pp. 505-535.

Gill, P.E. and W. Murray (1974) Newton-type methods for unconstrained and linearly constrained optimization, Math. Prog. 7, pp. 311-350.

Gill, P.E. and W. Murray (1977). The computation of Lagrange multiplier estimates for constrained minimization, National Physical Laboratory Rept. NAC 77.

Gill, P.E. and W. Murray (1979). The computation of Lagrange multiplier estimates for constrained minimization. Math. Prog., to appear.

Golub, G.H. (1965). Numerical methods for solving least squares problems, Num. Math. 7, pp. 206-216.

Hald, J. and K. Madsen (1978). A two-stage algorithm for minimax optimization, Inst. for Num. Anal. Reprot. NI-78-11, Tech. Univ. of Denmark.

Han, S.P. (1977a). A globally convergent method for nonlinear programming, JOTA 22, pp. 297-309.

Han. S.P. (1977b). Variable metric methods for minimizing a class of nondifferentiable functions, Dept. of Computer Science Rept., Cornell University.

Han. S.P. (1978a). Superlinear convergence of a minimax method, Dept. of Computer Science Rept., Cornell University.

Han, S.P. (1978b). On the validity of a nonlinear programming method for solving minimax problems, Math. Research Center Rept. 1891, University of Wisconsin, Madison.

Hettich, R. (1976). A Newton method for nonlinear Chebyshev approximation, in A. Dold and B. Eckmann, eds., Approximation Theory, Bonn.

Householder, A.S. (1964). The Theory of Matrices in Numerical Analysis, Ginn (Blaidsdell), Boston.

Jittorntrum, K. and M.R. Osborne (1979). Strong uniqueness and second-order convergence in nonlinear discrete approximation, working paper, Computing Research Group, Australian National University.

Kowalik, J. and M.R. Osborne (1968) Methods for unconstrained optimization problems, Elsevier, New York.

Madsen, K. (1975). An algorithm for the minimax solution of overdetermined systems of nonlinear equations, J. Inst. Math. Applics. 16, pp. 321-328.

More, J.J. and D.C. Sorenson (1979). On the use of idrections of negative curvature in a modified Newton method, Math. Prog. 16, pp. 1-20.

Murray, W. (1969a). Constrained optimization, Rept. MA 79, National Physical Laboratory.

Murray, W. (1969b). An algorithm for constrained optimization, in Fletcher, R. ed., Optimization, Academic Press, London and New York, pp. 247-258.

Murray, W. and M.L. Overton (1979). Steplength algorithms for minimizing a class of nondifferentiable functions, Computing 21, to appear.

Murray, W. and M.H. Wright (1978). Projected Lagrangian methods based on the trajectories of penalty and barrier functions, Systems Optimization Laboratory Rept. SOL-78-23, Stanford University.

Powell, M.J.D. (1977). the convergence of variable metric methods for nonlinearly constrained optimization calculations, Rept. DAMTP 77/NA3, University of Cambridge.

Rice, J.R. (1969). The Approximation of Functions, Vol. II, Addison-Wesley, Reading, Mass.

Robinson, S.M. (1974). Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms, Math. Prog. 7, pp. 1-16.

Rosen, J.B. and S. Suzuki (1965). Construction of nonlinear programming test problems, Comm. A.C.M. 8, p. 113.

Stewart, E.J. (1973). An iterative technique for absolute deviation curve fitting, J. Amer. Stat. Assoc. 56, pp. 359-362.

Swann, W.H. (1972). Direct search methods, in W. Murray, ed., Numerical Methods for Unconstrained Optimization, Academic Press, New York and London.

Watson, G.A. (1979). The minimax solution of an overdetermined system of nonlinear equations, J. Inst. Math. Applics. 23, pp. 167-180.

Wierzibicki, A.P. (1978). Lagrangian functions and nondifferentiable optimization, Rept. WP-78-63, Intl. Inst. Appl. Sys. Anal., Laxenburg, Austria.

Wilson, R.B. (1963). A simplicial algorithm for concave programming, Ph.D. thesis, Graduate School of Business Admin., Harvard University.

Wright, M.H. (1976). Numerical Methods for nonlinearly constrained optimization, Ph.D. thesis, Computer Science Dept., Rept. STAN-CS-76-566, Stanford University.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SOL 79-12	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR MINIMIX OPTIMIZATION		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Walter Murray and Michael L. Overton		8. CONTRACT OR GRANT NUMBER(s) DAAG29-79-C-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Research Office Box CM, Duke Station Durham, NC 27706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Operations Research Department - SOL Stanford University Stanford, CA 94305		12. REPORT DATE November 1979 13. NUMBER OF PAGES 67
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Optimization
Minimax
Quadratic programming

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOL 79-21 Walter Murray and Michael L. Overton

A PROJECTED LAGRANGIAN ALGORITHM FOR
NONLINEAR MINIMAX OPTIMIZATION

The minimax problem is an unconstrained optimization problem whose objective functions is not differentiable everywhere, and hence cannot be solved efficiently by standard techniques for unconstrained optimization. It is well known that the problem can be transformed into a nonlinearly constrained optimization problem with one extra variable, where the objective and constraint functions are continuously differentiable. this equivalent problem has special properties which are ignored if solved by a general-purpose constrained optimization method. The algorithm we present exploits the special structure of the equivalent problem. A direction of search is obtained at each iteration of the algorithm by solving an equality-constrained quadratic programming problem, related to one a projected Lagrangian method might use to solve the equivalent constrained optimization problem. Special Lagrangian multiplier estimates are used to form an approximation to the Hessian of the Lagrangian function, which appears in the quadratic program. Analytical Hessians, finite-differencing or quasi-Newton updating may be used in the approximation of this matrix. The resulting direction of search is guaranteed to be a descent direction for the minimax objective function. Under mild conditions the algorithms are locally quadratically convergent if analytical Hessians are used.